Haptic Feedback in Non-Euclidean VR Spaces:

A study with SenseGloves in Holonomy

A. Achilleos

Delft University of Technology

Haptic Feedback in Non-Euclidean VR Spaces

A study with SenseGloves in Holonomy

by

A. Achilleos

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday September 4, 2025 at 10:00 AM.

Student number:	4877837	
Project duration:	January 22, 2025 – September 4, 2025	
Thesis committee:	Prof. dr. M. Specht,	TU Delft, Thesis supervisor
	Dr. M. Skrodzki,	TU Delft, Daily supervisor
	Dr. R. Marroquim,	TU Delft, Committee member

Cover illustration by Sonia Balambanidou, created for this thesis

Abstract

Virtual Reality (VR) offers the possibility to explore and interact with complex digital worlds, yet natural locomotion is constrained by the limits of physical space. Hyperbolic geometry provides a compelling solution by embedding infinite virtual environments within finite areas, creating novel opportunities for research and design. This thesis investigates how embodied training and haptic feedback can enhance navigation and user experience in such non-Euclidean spaces. Twenty-eight participants took part in a between-subjects user study, using Holonomy VR, a hyperbolic VR application instrumented with the SenseGlove Nova 1 for force feedback and vibrotactile interaction. Participants were trained with either a drag-based embodied interface or a conventional button-based control scheme before completing matched navigation tasks. Performance was measured through speed, path efficiency and sequencing, while user experience was assessed through established questionnaires and interaction behaviour. The study finds that embodied training affords a practical advantage in subsequent navigation, and that perceived engagement with haptic elements is a strong predictor of positive usability, beyond the effects of task duration alone. Together, these results demonstrate that embodied practice and meaningful tactile interaction can help users adapt more effectively to non-Euclidean environments, offering both methodological contributions for VR research and design implications for the creation of more intuitive and engaging virtual worlds.

A. Achilleos
Delft, September 2025

Contents

1 Introduction

Virtual Reality (VR) has become an increasingly powerful medium, enabling rich and immersive interactions across a variety of fields such as entertainment, education, medical training, architectural visualisation, and scientific research. Its appeal lies in the ability to create experiences that closely replicate real-world interactions or, more intriguingly, allow exploration of entirely novel environments beyond physical constraints. Among these novel possibilities are environments built upon non-Euclidean geometries, like hyperbolic spaces, which have fundamentally different properties than the Euclidean geometry familiar from everyday experience.

Non-Euclidean geometries, particularly hyperbolic geometry, introduce fascinating properties that deviate significantly from everyday spatial intuitions, creating opportunities for unique spatial interactions and visualisations. This makes them particularly valuable for applications aiming to overcome physical limitations or intuitively visualise complex structures.

A key advantage of such a unique geometry, and why it is relevant in a navigational context for this thesis, is the inherent capacity to represent an infinite space in a compact form. This becomes especially useful in VR applications where physical boundaries are often a big restriction, requiring the use of movement illusions (teleportation, virtual locomotion, etc.) to navigate the virtual space. These unnatural movement mechanics are the main cause of (visually induced) motion sickness [9] in such applications due to the mismatch in the sense of motion between the visually induced one (in the VR application) and the sense of motion the user's body is experiencing in the real world (often standing still). This advantage can also be leveraged to visualise complex hierarchical structures more clearly, particularly in contexts where conventional Euclidean data representations would become overly dense or cluttered. Educational applications are also compelling, with the ability to offer a more intuitive understanding and engagement with complex mathematical concepts previously accessible only through abstract, formal instruction.

However, navigating these immersive virtual environments that defy Euclidean geometry poses a novel challenge for human spatial cognition. These VR spaces behave in counterintuitive ways that can obstruct and distort users' usual navigation strategies. Holonomy VR, discussed in more detail in Chapter 2, leverages the holonomy property to allow users to navigate an infinite virtual environment by walking, despite the physical confines. Having to use this property to navigate the hyperbolic scene makes it difficult for users to build accurate cognitive maps of the scene.

Prior work suggests that, despite these challenges, humans *can* adapt to and navigate non-Euclidean spaces with the right support. A recent study by Pisani et al. had participants navigate analogous tasks in Euclidean vs. hyperbolic VR levels. Interestingly, users did *not* become highly disoriented in the hyperbolic condition and even found certain complex structures less confusing to navigate than in Euclidean space [33]. With minimal training, several users reported they "had not realised it was an unusual space" when first experiencing the hyperbolic world. Holonomy VR allows participants to physically walk within a tracked 3×3m area, using natural locomotion as the input for virtual movement

[44]. Such embodied interaction made movement control intuitive and, notably, aided in preventing motion sickness. This aligns with broader VR research showing that physical locomotion reduces sensory conflict and improves user comfort compared to joystick or teleportation locomotion [38] [26]. However, physical free-walking interfaces are constrained by real-world space and may not, by themselves, teach users the underlying spatial relationships of the warped world. There is a need for methods that both accelerate spatial learning of non-Euclidean layouts and enhance user experience (usability, comfort, immersion), without requiring large tracking areas or costly hardware setups.

This thesis investigates a dual approach to improving user performance and experience in a hyperbolic VR environment (1), an *embodied learning* intervention to foster spatial understanding, and (2) the use of *haptic feedback hardware* to provide a more natural interaction during navigation. The central idea is that engaging users' sensorimotor abilities during training/acclimation will help them form better mental models of the strange geometry properties and feel more ”at home” when navigating. Embodied cognition theory stipulates that cognitive processes (like spatial reasoning) are deeply grounded in the body's perceptual and motor experiences [50]. In line with this, physical movement and interaction are known to be critical for spatial learning [26]. For example, Ruddle et al. [38] found that allowing participants to physically move (walk or turn) in VR resulted in significantly more efficient navigation and spatial memory than purely using joystick controls. It is hypothesised that applying these principles in a targeted training phase will improve users' ability to handle hyperbolic navigation. In this intervention, one group of participants (treatment) will learn the Holonomy movement mechanics via drag-based embodied controls to grab and move a pawn within a 9-tile grid, while a control group learns via traditional button-based controls. Both groups make use of haptic feedback hardware. The study design is detailed further in Chapter 5. This kind of embodied learning approach has shown promise in other domains; for instance, VR studies in education report that manipulating virtual objects with natural hand movements can ground understanding of abstract concepts in sensorimotor experience [8]. This research aims to test whether such benefits carry over to spatial skills in a non-Euclidean VR context.

The second facet of this thesis addresses the user interface and feedback during the main VR navigation tasks. In standard VR setups, users usually interact by using hand controllers, which only provide limited vibrotactile feedback and use abstract input (button presses or joystick tilts) to represent actions like grabbing or moving. Replacing the controllers with haptic gloves introduces a far more natural mapping between real and virtual actions. The SenseGlove Nova 1 device used in this study is a wireless force-feedback glove that tracks the user's finger movements and provides kinaesthetic and tactile feedback for virtual interactions [41]. With the SenseGlove, a user can literally grasp at virtual objects or surfaces and feel resistance to simulate touch and weight. More information about the SenseGlove follows in Section 2.3. Such rich feedback and one-to-one hand mapping are expected to heighten immersion and presence. Indeed, recent research comparing such gloves to standard controllers found that gloves yielded a significantly greater sense of spatial presence and embodiment, as users felt their natural hand movements were directly reflected in VR. [31]. Participants using gloves describe the interaction as ”highly natural”, leading to deeper cognitive absorption in the virtual task. In a dense and potentially disorienting environment like Holonomy VR, it is anticipated that the gloves will improve usability (by means of more intuitive interactions and reduced mental effort) and comfort (by mitigating the disconnect between action and virtual feedback). It is important to note that these assumptions need to be empirically verified. Some studies have noted that adding haptic feedback does not improve objective task performance or spatial learning, potentially due to increased cognitive load or device encumbrance [40].

In summary, this research tackles a fundamental question: *Can we leverage embodied learning techniques and haptic technology to improve how people navigate and experience non-Euclidean virtual worlds?*. To this end, an experiment is conducted in the Holonomy hyperbolic VR environment with a controlled intervention in the training stage (embodied vs. conventional learning). The experiment is built to exaggerate the use of haptic feedback hardware and its impact on user experience. The study is designed to shed light on both performance outcomes (e.g., navigation efficiency, error rates, spatial learning transfer) and user experience outcomes (e.g., usability ratings, (dis)comfort, sense of presence).

2
Background

This chapter provides foundational knowledge required to understand the concepts and techniques explored in this thesis. Hyperbolic geometry and its differences from Euclidean geometry are introduced, followed by the Holonomy VR environment that this research builds upon. Lastly, an overview of haptic technology, particularly gloves, is given, alongside its role for interactions in VR spaces.

2.1. Hyperbolic Geometry

Euclidean geometry, formalised by the ancient mathematician Euclid, is founded on five key postulates. The fifth postulate states that if a line segment intersects two straight lines, which form two interior angles on the same side that are less than 90°, then if the lines are extended indefinitely, they will intersect. This can be visualised better in Figure 2.3. Non-Euclidean geometries challenge the fifth postulate, commonly known as the parallel postulate, and propose alternative axioms. Two prominent types of non-Euclidean geometries stem from this: elliptic and hyperbolic; this thesis focuses on the latter, the same geometry also being used in Holonomy VR. However, elliptic geometry provides a noteworthy example to understand the parallel transport effect. Their main differences can be seen in Figure 2.1 below, where the same triangle and circles are plotted on the surface of each geometry.

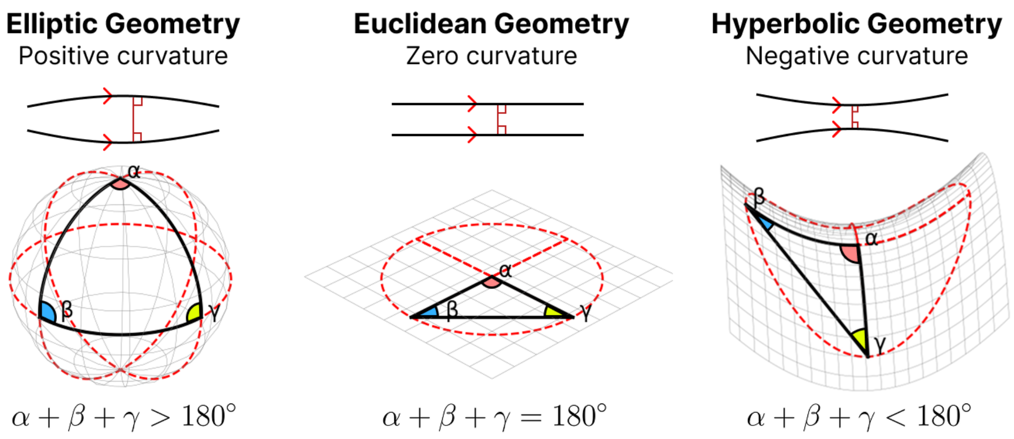


Figure 2.1: A right-angle triangle drawn with each geometry and their main differences illustrated.

Consider the elliptic geometry that examines principles on the surface of a sphere. Parallel transport involves moving a vector along such a curved path while maintaining its orientation relative to the surface's curvature. Parallel transport in such a (curved) geometry can alter the vector's orientation, an effect not observed in flat, Euclidean spaces. Figure 2.2 shows an illustration of this in elliptic geometry, and an intuitive real-world example of parallel transport could be given as follows: An arrow-shaped cart is placed on the North Pole (A), pointing east, and its starting orientation is marked on the ground. The cart is never rotated and is dragged sideways from the North Pole to the equator (B). From there, it follows the equator line being pushed forward until a quarter of the equator's length is covered (C). From that point, the cart is then dragged sideways back to the North Pole (A). The cart will now be rotated 90° relative to its original marked orientation. This direction change is precisely what holonomy is, and is a result of parallel transport along a curved surface such as the Earth. Holonomy VR makes use of this effect, hence the name of the project, since it also exists in hyperbolic geometry.

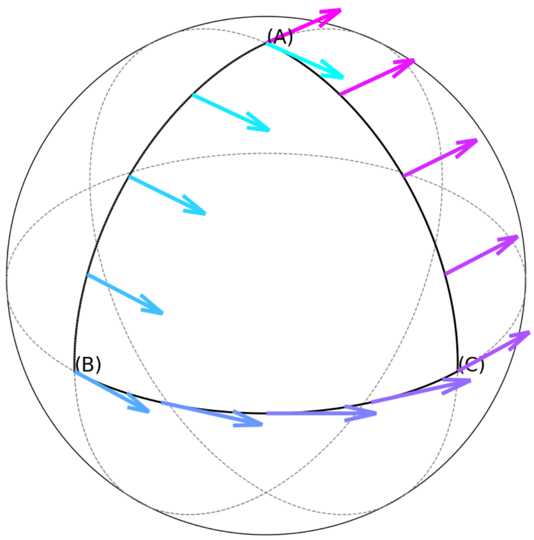


Figure 2.2: Parallel transport following A→B→C→A. The vector is rotated 90° counter-clockwise at the starting position.

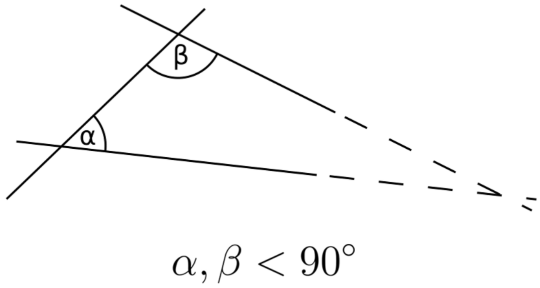


Figure 2.3: Euclid's fifth postulate visualised. With $\alpha, \beta < 90°$, the line segments will intersect if extended. [16]

Hyperbolic geometry replaces this fifth postulate with a rule stating that through any given point not on a given line, infinitely many parallel lines can be drawn, none of which intersect with the original line or each other [43]. This deviation creates fascinating properties such as the angles in triangles summing to less than 180°, parallel lines diverging from one another, and most notably, the space expands exponentially as one moves away from any given point. This expansion and the holonomy effect are the enabling properties for Holonomy VR to represent compactly and allow for exploration of an infinite area even in a constrained physical space. This is explained in more detail in the following section.

Navigating virtual environments constructed on hyperbolic geometry introduces unique cognitive challenges. The unfamiliar curvature and holonomy effects disrupt traditional spatial strategies, making it difficult for users to form reliable mental maps or predict how their movement will affect their orientation. In Holonomy VR, users often experience disorientation when first encountering these conditions, especially when the environment appears to shift unexpectedly during looped movements. This thesis directly addresses these challenges by investigating whether haptic feedback through wearable glove-based hardware can support users in adapting to the hyperbolic space and improving their experience. This is discussed in more detail in Section 4.2.

2.2. Holonomy: A Virtual Reality Environment in Hyperbolic Space

Holonomy VR is a "non-Euclidean labyrinth" game set on a model of a hyperbolic world [46] designed in Unity to facilitate exploration of hyperbolic geometry. It currently supports three game modes: finding a flag, finding keys to open a chest, and finding landmarks. For the purpose of this thesis, this will be expanded upon to facilitate haptic feedback interaction. Holonomy VR provides users with an infinitely large, explorable hyperbolic world through natural locomotion, allowing them to physically walk within a confined real-world space (typically around a 3×3 meter area), eliminating the need for artificial VR movement methods such as teleportation or joystick movements [44]. This intuitive approach significantly enhances immersion, reduces VR-induced motion sickness, and closely aligns virtual navigation with physical motion.

The application leverages a hyperbolic square tiling system where five squares meet at each vertex, creating the smallest feasible hyperbolic configuration that still exhibits essential hyperbolic properties, notably the phenomenon of holonomy itself. The environment is constructed from square tiles arranged in a 3×3 grid, bounded by a hedge to prevent movement and viewing tiles beyond this area. This setup creates the illusion of a confined play space, where users are seemingly limited to 2 steps in any direction. Yet, because of the hyperbolic structure, the environment can incorporate more than the visible 9 tiles. Unlike Euclidean space, where at most 4 squares meet at a single vertex, the order-5 square tiling, see Figure 2.4, constructs a hyperbolic plane with more tiles existing in the 9-tile grid. In VR, these tiles are hidden unless a user stands and orients themselves in a position where a fifth tile is shown around a particular point, see Figure 2.5 for an example. This tiling structure enables spatial compression but causes a misalignment between physical and virtual coordinates, making it difficult for users to navigate in the space. Users see, move, and exist in a seemingly Euclidean play area, but that is merely a "truncated" projection of a larger hyperbolic world centred on their current location. To shift this grid forward, for example, users have to use the holonomy property to reveal new tiles by completing a loop.

One of the ways to model a hyperbolic plane is the Poincaré Disk Model, which in Holonomy VR is used as the minimap, see Figure 2.6, and is a core component of the navigation system. The hyperbolic plane is mapped inside a unit circle, and its circumference border is set at infinity, with distances growing exponentially as previously mentioned. It makes up a compact and visually comprehensible depiction of hyperbolic geometry in two dimensions. The minimap aids navigation by offering visual cues about the user's position and the surrounding environment, although interpreting these cues effectively can still be challenging due to the non-intuitive geometric relationships of the hyperbolic space and the real Euclidean world users themselves exist in.

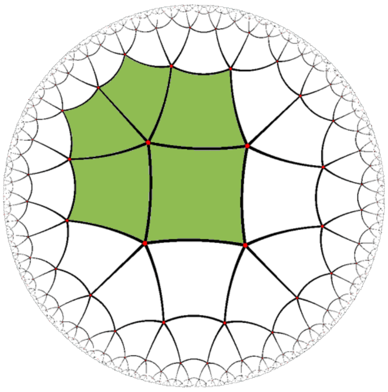


Figure 2.4: Order 5 hyperbolic tiling with 5 squares highlighted around a single vertex. [6]

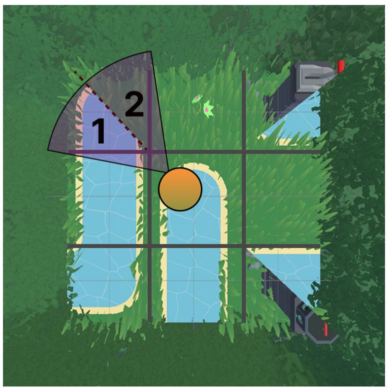


Figure 2.5: The confined top-down view of the play area with the tile grid overlaid. Tiles 1 and 2 are both visible from the indicated location.

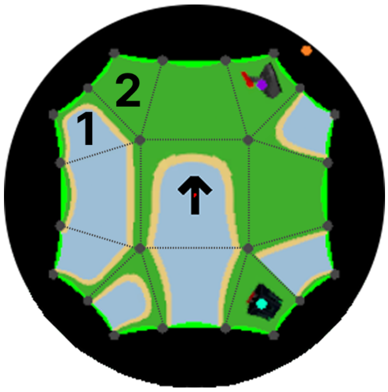


Figure 2.6: The minimap corresponding to the top down view in Figure 2.5. Tiles 1 and 2 are noted as seen on the minimap, modelling the hyperbolic environment.

Users frequently report initial confusion and disorientation as they attempt to make sense of the unfamiliar spatial dynamics. The environment's compactness and shifting geometry often elicit sensations of claustrophobia or spatial discomfort during early interactions. Forming accurate cognitive maps is particularly difficult, as users must reconcile global inconsistencies with their intuitive Euclidean expectations. Consequently, successful navigation demands significant acclimation and the development of new spatial cognition strategies tailored to the non-Euclidean context. Holonomy's unique implementation of hyperbolic geometry and embodied locomotion thus provides a compelling platform to explore whether enhanced sensory feedback, specifically haptics as explained in the next section, can help mitigate these challenges.

2.3. Haptics

Haptic technology refers to the systems that simulate the sense of touch by applying forces, vibrations, or motions to the user, enhancing interaction by engaging the tactile modality. In the context of virtual reality, haptics serve as a critical complement to visual and auditory feedback, enabling more immersive, intuitive, and embodied experiences. By incorporating the sense of touch, haptics reduce the abstraction inherent in traditional VR control schemes, such as handheld controllers, and allow users to interact with virtual environments using familiar motor patterns.

Among the various types of haptic interfaces, glove-based systems have gained attention for their ability to replicate the fine motor control of human hands. Unlike standard VR controllers, which offer limited vibrotactile feedback through a few contact points, haptic gloves afford more granular, spatially distributed feedback across the hand and fingers. This natural mapping of interaction, where users can touch, grasp, and manipulate virtual objects with gestures that mirror real-world behaviour, can significantly enhance presence and reduce cognitive load.

In this thesis, the SenseGlove Nova 1 is employed as the haptic interface of choice [41]. The Nova 1 is a lightweight, wireless, exoskeleton-style glove designed specifically for VR and AR applications. Two primary forms of feedback are offered:

- **Force-feedback:** Achieved through mechanically actuated tendons that restrict finger motion to simulate resistance or object solidity during grasping, with up to 20 N of force.
- **Vibrotactile-feedback:** Delivered through actuators placed on the thumb, index, and palm. Enables users to feel impacts, contact events, or any type of notification conveyed through vibration.

This combination allows users to experience the sensation of physically interacting with the virtual world. For instance, when a user's hand collides with a virtual object in Holonomy VR, the glove can simultaneously stop finger movement to simulate resistance and emit a vibration to represent the moment of contact. Such feedback can deepen immersion, improve spatial awareness, and reinforce the link between motor actions and virtual consequences.

The integration of the SenseGlove into Holonomy VR is particularly relevant given the spatial disorientation and perceptual challenges introduced by the non-Euclidean environment. Tactile feedback may help reduce feelings of detachment or confusion by anchoring the user's experience through physical sensation. Research shows that the combination of vision and proprioception significantly improves spatial orientation and memory for self-motion in immersive VR [4], underscoring the potential of haptic gloves to provide the stabilising sensorimotor cues needed in such disorienting contexts. It also offers an opportunity to investigate how enhanced embodiment, via hands-on interaction with the virtual world, can improve users' ability to understand and navigate hyperbolic spaces.

In this thesis, the SenseGlove Nova 1 serves both as a tool for embodied training in the intervention phase and as an input device in the main Holonomy task. Its role is not only functional but investigational: to determine whether haptic-enhanced interaction leads to better navigation outcomes and higher overall user experience in non-Euclidean virtual environments.

3 Related Work

This chapter reviews relevant literature on embodied learning in virtual reality, haptic feedback technologies, and spatial navigation in non-Euclidean VR environments. Firstly, the theoretical foundations of embodied interaction and embodied learning in VR are discussed, drawing on recent frameworks that explain how physical action and context can ground abstract understanding in immersive settings. Prior work is then examined, related to spatial orientation and navigation in virtual environments, with a focus on the unique challenges posed by non-Euclidean geometries such as hyperbolic or curved virtual spaces. Finally, research on haptic feedback in VR is reviewed, particularly the use of glove-based interfaces, and how tactile cues can enhance user interaction, presence, and learning. Highlighted throughout this chapter is the rationale for combining an embodied training approach with a haptic interface (the SenseGlove Nova 1) to improve users' spatial learning and experience in the Holonomy hyperbolic VR application.

3.1. Embodied Interaction and Learning in VR

Virtual reality offers the opportunity for embodied learning, where learners actively use their bodies to interact with digital content, potentially leading to deeper understanding. The notion of *embodied cognition* posits that cognitive processes are fundamentally grounded in the body's sensorimotor experiences [50]. In other words, how we think and learn is tightly coupled with how we move and perceive through our bodies. This contrasts with traditional "disembodied" views that treat the body as just an output device for the brain's commands.

3.1.1. Embodied Learning in Educational Contexts

Alibali et al. identify three ways in which gestures can be used in teaching and learning, and argue that mathematical cognition is embodied in both perception and action [3]. Hostetter and Alibali propose that gestures stem from mental simulations and serve to facilitate the retrieval of imagery and embodied representations of language and knowledge [18]. This embodied perspective has influenced research in Virtual Reality (VR) and Mixed Reality (MR). For instance, Lindgren and Johnson-Glenberg base their view of embodied learning on the premise that bodily actions can enhance cognitive processes, such as language and memory retrieval. This way, increasing an individual's range of meaningful physical experiences may lay the groundwork for new conceptual structures [28]. Supporting this, Johnson-Glenberg et al. found that students who learned STEM concepts in immersive, body-engaging MR environments demonstrated significantly greater understanding than those receiving traditional instruction [24], suggesting that linking ideas to sensorimotor activity can enhance conceptual learning. Similar implications emerged in a study by Prakash and Rajendran [35], who implemented interaction behavioural data logging in a VR learning environment. Their analysis of learners' embodied actions revealed patterns linked to deeper engagement and suggested the potential for improved learning outcomes through embodied interaction.

3.1.2. Frameworks for Embodied Interaction Design

Embodied interaction is a key component of this approach, referring to the design of interactive systems that involve bodily actions to manipulate virtual objects or environments. Chatain et al. describe a theoretical framework for embodied learning in VR in which both the *physical* and *social* context influence the learner's bodily interactions [8]. According to this framework, meaningful learning emerges from goal-directed physical actions that are situated within a context, for instance, manipulating virtual objects with one's hand in a way that aligns with real-world experiences, ideally in a social or environmental setting that supports the learning goals. In their aforementioned framework, the authors highlight three perspectives for embodied learning in VR as can be seen in Figure 3.1 below:

- **Embodied cognition:** The idea that thinking is grounded in bodily experiences, physical actions can support conceptual understanding and memory formation.
- **Embodied interaction:** Emphasises the goal-directed engagement with virtual objects and environments, shaped by the physical and social context of the user.
- **Avatar embodiment:** Captures the extent to which users identify with their virtual body, and how bodily ownership and agency contribute to learning and engagement.

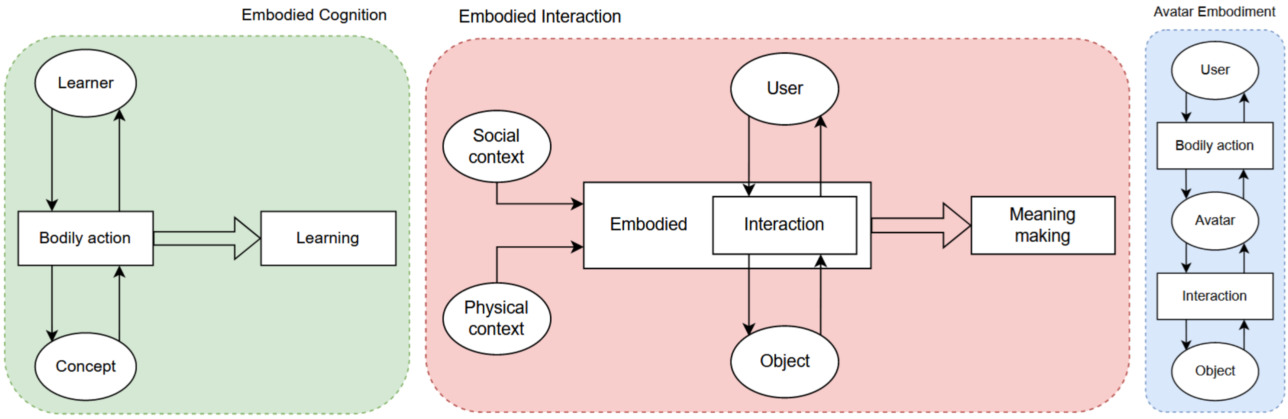


Figure 3.1: Representation of the three perspectives proposed by Chatain et al. in their work. Ellipses represent actors, and rectangles represent processes. The relation A → B means that A *informs* B, while C ⇒ D means that C *induces* D. All arrows are conditional, e.g., a certain process may happen but does not necessarily. [8]

A well-designed VR learning activity ideally harmonises these aspects. Figure 3.2 illustrates the proposed framework by Chatain et al., and how these perspectives interrelate. Bodily actions informed by a learner's context and mediated through an avatar can lead to meaning-making and learning outcomes. Important to note, these perspectives do not automatically coincide; the authors formulate conditional relationships, but they are not guaranteed without support from a congruent design. For example, high embodied interaction does not warrant embodied cognition benefits unless the interactions are thoughtfully aligned with the learning objectives. Similarly, a realistic avatar embodiment can sometimes be broken in favour of learning; Chatain et al. found that even non-intuitive avatar designs can support meaning-making if they direct the user's focus to key concepts. These insights underscore the need for intentional design strategies that align bodily activity with learning objectives.

In practical terms, this framework justifies the use of physical interaction to support abstract concept acquisition. Prior work confirms that such a design can improve learning outcomes: when learners can grab, gesture, or move as part of the lesson, they develop sensorimotor links to the content, making it more memorable and intuitive [24]. In this research, this model provides the foundation for introducing an embodied training phase, where users explore the logic of hyperbolic movement through haptic-supported interaction, before entering the full Holonomy VR environment. The goal is to design the intervention to activate all three layers of embodiment, a deeper and more intuitive understanding of the spatial structure and its navigation in a non-Euclidean world.

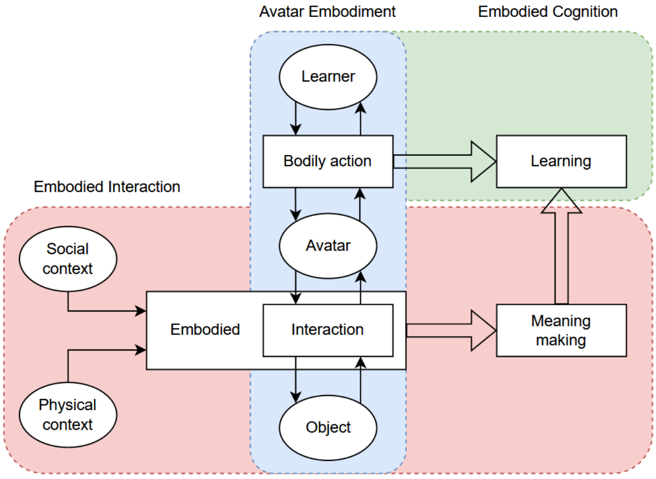


Figure 3.2: The framework proposed by Chatain et al. in their work combines all three perspectives for embodied learning. Ellipses represent actors, and rectangles represent processes. The relation A → B means that A *informs* B, while C ⇒ D means that C *induces* D. All arrows are conditional, e.g., a certain process may happen but does not necessarily. [8]

3.2. Spatial Navigation in Virtual Environments

Navigating virtual environments (VEs) directly impacts spatial learning, cognitive engagement, and user comfort. Over the years, various locomotion techniques have been explored, each with distinct benefits and drawbacks regarding user immersion, spatial cognition, cognitive load, and motion sickness. This section reviews key findings on how different navigation methods influence spatial cognition and performance, contrasting embodied (physical) approaches and abstract (controller-based) approaches across both non-Euclidean and Euclidean virtual spaces.

3.2.1. Natural Locomotion and Embodied Navigation

Embodied interaction through natural locomotion, specifically physical walking and turning, has been shown to significantly improve navigational search performance in virtual environments. Ruddle and Lessels [38] demonstrated that participants who physically walked and turned in an immersive VR setting completed a spatial search task with far greater efficiency than those limited to visual input or rotational movement alone. Their results highlight the critical role of the ability to move one's body naturally (even within a limited tracking area) provides cues that complement visual information, leading to better formation of cognitive maps of the virtual space

Early evidence also came from Pausch et al. [32], who quantified the effects of immersive VR on visual search tasks. While immersive VR did not significantly improve users' ability to find targets in camouflaged scenes, it enabled users to more efficiently determine when no target was present, likely due to a better internal model of the space. Notably, training in VR transferred positively to desktop use, while the reverse led to performance degradation. A landmark study by Usoh et al. [45] compared three locomotion modes: actual walking, walking-in-place, and virtual flying. They reported a clear hierarchy of presence and spatial engagement: *"Walking > walking in-place > flying"*, and found that presence was strongly correlated with users' sense of embodiment in their avatars. This finding was further supported by Ruddle et al. [39], who demonstrated that translational body-based cues, specifically walking, are critical for building accurate cognitive maps in virtual environments, especially when these environments are both large in scale and extent.

Aligning with current evidence, Waller and Hodgson [47] highlight that full-body translational movement provides body-based sensory inputs that support spatial tasks like orientation and path integration. In contrast, stationary navigation using a controller deprives users of these cues, often creating sensory conflicts when vision indicates movement but the body senses none. Nguyen-Vo et al. [30] in their 'NaviBoard' and 'NaviChair' experiments demonstrated that full physical rotation paired with limited translational cues through leaning or stepping significantly improves virtual locomotion performance and user experience. These interfaces nearly matched physical walking and outperformed joystick-based navigation by a wide margin. Similarly, Reuterswärd [37] found that combining vertical locomotion with embodied interaction in 'impossible spaces' significantly enhanced user immersion across spatial, emotional, cognitive, and tactical dimensions.

The gathered evidence and support from literature bode well for the setup in Holonomy VR, where natural locomotion is the method of traversing the VR space. It must be noted, as previously mentioned, that embodied interfaces do not universally guarantee better learning; their advantage can depend on task complexity and user strategy. While embodied interaction often provides qualitative benefits (greater presence, engagement, attention allocation), its impact on measurable learning outcomes may vary with context. Researchers advocate examining *how* learners' bodies are engaged during VR training and aligning design with learning objectives.

3.2.2. Abstract and Artificial Locomotion Techniques

While natural locomotion provides significant cognitive and experiential advantages, practical constraints often warrant alternative locomotion methods such as the aforementioned joystick controls, teleportation, or even thought-based navigation. These abstract locomotion methods differ significantly in their cognitive and experiential impacts.

Teleportation, though effective in reducing cybersickness compared to smooth locomotion methods, may compromise users' spatial learning and spatial updating. Cherep et al. [10] demonstrated that teleportation disrupts continuous spatial awareness due to the fragmented view of the environment, leading to greater disorientation and navigation errors. Similarly, Christou and Aristidou [11] found that although teleportation yielded the lowest levels of cybersickness among tested locomotion methods, it often resulted in users overlooking spatial details and collecting fewer environmental cues, indicating diminished spatial engagement. Langbehn et al. [27] reinforced these concerns by showing that teleportation led to significantly poorer spatial orientation performance and increased angular errors compared to real walking and redirected walking. In contrast, thought-based navigation, as explored by Friedman et al. [13], introduces a novel interaction paradigm via motor imagery and brain-computer interfaces (BCIs), offering hands-free control in immersive environments. However, their study highlighted that while immersive displays enhanced presence and BCI accuracy, users encountered substantial cognitive and usability challenges due to the indirect and abstract mappings between mental commands and virtual movement.

Kimura et al. [25] found that while participants were able to navigate and reorient in both real-world and immersive virtual environments, their reliance on spatial cues differed. In particular, geometric cues were encoded less accurately in VR, leading participants to rely more on encoded feature information compared to those in real-world conditions. This hints towards the fact that spatial cue use in VR may not fully replicate real-world navigation behaviour, even in highly immersive setups.

3.2.3. Spatial Navigation in Non-Euclidean Virtual Environments

Navigating non-Euclidean spaces such as hyperbolic or curved environments poses unique cognitive challenges, mainly because human spatial cognition appears strongly biased toward Euclidean assumptions. Studies by Widdowson and Wang [48][49] demonstrate that participants persistently relied on Euclidean mental models even in environments with clear, visually curved, non-Euclidean geometry. Their findings indicate that people systematically default to Euclidean spatial updating and representations, regardless of perceptual evidence to the contrary. This urges the need for targeted interventions to facilitate adaptation to non-Euclidean environments and steer away from these deeply ingrained intuitions.

However, humans are not all helpless in non-Euclidean spaces; Pisani et al. [33] found that users were able to adapt well to navigation in hyperbolic VR spaces, indicating that non-Euclidean geometries are not inherently disorienting in immersive environments. In fact, users sometimes reported that navigating branching structures like trees felt more intuitive in hyperbolic space. Similarly, Jaksties et al. [23], in their study of orientation within *Magical Tower VR*, observed that players successfully navigated a non-Euclidean level structure and responded positively to the experience, suggesting that with appropriate design elements, such as landmarks and spatial consistency, effective navigation is achievable in non-Euclidean VR worlds.

An additional practical example of navigating hyperbolic environments is seen in the VR game *Hyperbolica*[20]. In this first-person adventure, players find themselves in a "reality-warping" hyperbolic world, where straight lines diverge, space grows exponentially, and the horizon is curved. Although its geometry can be initially disorienting, both developers and players report that users quickly built an intuitive sense of navigation through exposure to non-Euclidean mechanics and use of familiar landmarks. Another notable example is *HyperRogue* [21], a rogue-like game set on the hyperbolic plane that offers a non-VR but conceptually related experience. In gameplay, tiles shrink toward the edges due to hyperbolic mapping; fog-of-war ensures the world feels vast and inexhaustible. Players not only adapt to but also exploit the geometry to their advantage. Together, these two examples demonstrate how a well-designed experience in hyperbolic spaces allows humans to adapt meaningfully to spatial distortions, supporting the argument that non-Euclidean VR navigation is both feasible and engaging.

3.2.4. Enhancing Spatial Understanding Through Embodied Pre-Training

Given the inherent complexity of non-Euclidean navigation, preliminary embodied training may significantly benefit users by establishing initial sensorimotor familiarity before exposure to a complicated virtual environment. Abrahamson and Sánchez-García [1] articulate how physical interaction and embodied cognition theories support the idea that embodied experiences facilitate comprehension of mathematical concepts. The authors argue that learning mathematics is not just cognitive or symbolic, but deeply sensorimotor. They found that learners first executed random hand movements and, over time, began to coordinate their motions. Eventually, the learners discovered the mathematical ratio as an emergent property of their sensorimotor coordination.

Alibali and Nathan [3], as previously stated, provide empirical evidence that embodied interactions such as gestures significantly enhance abstract reasoning and convey structure, order, and relationships between concepts. The aforementioned embodied interaction framework proposed by Chatain et al. further supports structured, preliminary embodied training as crucial for aligning sensorimotor experiences with subsequent cognitive tasks.

In the context of this Holonomy VR study, the embodied pre-training approach (explained in more detail in Subsection 5.4.1) leverages these insights. By using drag-based controls to physically manipulate a simplified representation, participants activate cognitive and sensorimotor systems relevant to hyperbolic navigation, potentially facilitating improved learning outcomes within the complex VR environment.

3.2.5. Comparing Navigation Interfaces: Drag-based vs. Button-based Controls

A critical component of this research is the comparative evaluation of embodied drag-based controls versus abstract button-based navigation methods. Recent studies have explicitly compared these interface types, highlighting nuanced differences in user experience and performance.

Huang et al. [19] examined different levels of embodied interaction in a study involving navigation and analysis of network visualisations across four interfaces: standard 2D visualisation with mouse input, 3D visualisation with mouse, 3D visualisation with physical trackball, and immersive VR with handheld controllers. Their findings indicated that while the 3D immersive VR interface significantly enhanced accuracy and reduced perceived workload in tasks involving spatial exploration (counting triangles within a network), it performed poorly in tasks requiring precise comparative analysis due to perspective distortions and increased cognitive load. Thus, embodied VR interactions demonstrated task-dependent benefits, suggesting that highly embodied interfaces are most effective for exploratory tasks requiring spatial understanding, while simpler interfaces remain advantageous for tasks requiring quick comparative judgments or precision.

Similarly, Bektaş et al. [5] developed and evaluated the Limbic Chair, an embodied control interface that uses leg-based movements to control avatar locomotion in VR, comparing it to traditional gamepad-based controls across city navigation and flight simulation tasks. While gamepad controls led to faster task completion and were rated higher in usability, likely due to user familiarity, the Limbic Chair offered reduced simulator sickness and was described as more immersive and natural by some participants. The study concluded that embodied interfaces like the Limbic Chair can enhance user experience in certain VR contexts, despite trade-offs in workload and performance efficiency."

These findings underline the importance of considering both the nature of the task and user familiarity when designing VR navigation interfaces. However, the extent to which these embodied interaction advantages extend robustly into non-Euclidean contexts remains under-investigated. This research aims to further investigate these findings by comparing drag-based and button-based navigation controls within the context of complex spatial tasks in virtual reality. By employing both control interfaces utilising the SenseGlove to maintain hardware consistency, it can be shown how different interaction mappings influence cognitive load, spatial understanding, and user experience. This comparative approach directly addresses critical gaps identified in current literature and aims to inform best practices for interface design in complex spatial navigation contexts.

3.3. Haptic Feedback and Natural Interaction in VR

Haptic feedback technologies are a key enabler of embodied interaction in VR, allowing users to physically feel and manipulate virtual objects through tactile and force-based stimuli. This section synthesises current knowledge on haptic modalities, their applications in education and spatial learning, and the implications for user experience and interface design.

3.3.1. Overview of Haptic Interfaces and Capabilities

Haptic feedback in VR is typically categorised into three modalities:

- **Tactile:** Surface-level vibration or pressure
- **Kinaesthetic:** Force-feedback that resists motion
- **Proprioceptive:** Body positioning or balance-related

The SenseGlove Nova 1 used in this research utilises all three modalities, as explained in Section 2.3, with the third being achieved by taking advantage of the VR controllers mounted on the gloves and used in conjunction with an HMD. Shi and Shen [42] provide a comprehensive taxonomy of these techniques, emphasising their relevance to immersive environments. Their review categorises tactile sensing by mechanism and haptic feedback by actuator type. These modalities enable a broad range of applications, from gesture recognition and virtual object manipulation to medical rehabilitation and remote robotic control. However, the authors warn that more realistic feedback often requires bulky hardware, with latency and energy efficiency being crucially important. High-fidelity haptics are still expensive and not easily mass-produced, and the next generation of haptic VR systems calls for interdisciplinary collaboration between engineering, biomechanics, and material science fields.

Smart gloves have emerged as one of the most promising tools for integrating haptic feedback into VR. According to Caeiro-Rodríguez et al. [7], commercial gloves, such as the SenseGlove Nova 1 used in this research, offer multiple degrees of freedom and force feedback, enabling realistic hand-object interactions. In their review of 24 commercially available gloves, the authors found that most gloves focused on tracking, with far fewer providing haptic feedback. Integration of such devices in any system is often complicated, with limited products offering standardisation and interoperability. Similar limitations are identified here; gloves are often developed for niche or specialised cases, with comfort, ergonomics, and battery life being major design trade-offs.

Irigoyen et al. [22] highlight the broad utility of haptic technologies, not limited to glove-type devices, in educational, rehabilitation, and training settings. They emphasise these systems' capacity to enhance user engagement and support skill acquisition, particularly in medical education and specialised learning contexts, through real-time feedback. Their review suggests that combining visual and haptic cues can improve immersion and task comprehension, although with similar technical and adoption barriers as previously mentioned.

In a study by Afzaal and Alim [2], a haptic-enabled stylus was compared with traditional VR handheld controllers and found that interaction performance varied by task. The stylus enabled faster and sometimes more accurate point localisation, especially when users physically contacted the surface, while handheld controllers allowed faster curve brushing. Nonetheless, they concluded there is no universally superior device, interaction effectiveness is task-specific, and system designers should consider task requirements and user ergonomics when choosing input modalities for VR environments.

Palombo et al. [31] qualitatively investigated how VR gloves compare to handheld controllers in influencing spatial presence, embodiment, and cognitive absorption. In a crane assembly task involving 20 participants, the VR glove was consistently rated as more natural and immersive than the HTC Vive and Valve Index controllers. Participants reported deeper engagement and a stronger sense of ownership over the virtual hand. While no performance metrics were collected, the study highlights how naturally mapped input devices can enhance subjective VR experiences, suggesting experiential benefits even in the absence of haptic feedback.

Earlier work by Poorten et al. and van der Meijden and Schijven [34][29] offers a nuanced view of haptic feedback in surgical training. Poorten et al. reviewed the literature and found mixed evidence for the value of haptic feedback, especially noting concerns over the quality and appropriateness of haptic implementation in VR simulators. Van der Meijden and Schijven similarly concluded that while haptic feedback shows promise, particularly in robot-assisted and complex surgical tasks, empirical support remains inconsistent and context-dependent

3.3.2. Haptic Feedback for Learning and Spatial Cognition

Several studies have examined how haptic interaction influences cognitive performance in spatial tasks. Ruthenbeck et al. [40] investigated whether visuo-haptic interaction in a VR nasendoscopy simulation would enhance spatial learning compared to visual-only exploration. Contrary to intuitive expectations, their study found no significant improvement in recall accuracy or response times for the haptic group. The authors suggest that the added cognitive load of manipulating an unfamiliar haptic device may have interfered with spatial encoding, particularly for novice users. This implies that, unlike in motor-intensive tasks, haptic feedback may not offer a learning advantage for spatial cognition, especially when the interface itself demands considerable mental effort.

Similarly, Forgiarini et al. [12] compared force-feedback gloves with standard handheld controllers in a VR object-location memory task. Despite users reporting a higher sense of presence with the gloves, the study found no significant differences in memory accuracy between the two groups. In fact, participants using the gloves completed tasks more slowly and reported greater physical fatigue, particularly in the fingers and wrists. The authors suggest that the added physical demands of force-feedback gloves may limit their effectiveness for spatial memory training in this context. As with Ruthenbeck et al.'s results, the study suggests that haptic feedback, while engaging, may not enhance spatial memory and should be deployed sensibly in cognitive training contexts.

The previously mentioned study by Palombo et al. [31] found that VR gloves, due to their realistic natural mapping, led to higher levels of cognitive absorption and embodiment. Complementing these findings, Gibbs et al. [15] demonstrated that the integration of haptic and visual feedback in VR environments significantly enhanced users' sense of presence. In their experiment, participants interacted with a virtual stick as a ball bounced across it under varying feedback conditions. Presence scores were consistently highest when both modalities were combined. These results underscore the importance of multimodal sensorimotor sensations in creating convincing and immersive virtual experiences. Tactile feedback plays a critical role in how users engage with and make sense of virtual environments.

Gao et al. [14] investigated the use of bi-manual haptic feedback, delivered through standard VR controllers, for supporting spatial search tasks in virtual reality. They found this approach significantly improved the users' ability to recognise spatial directions with high accuracy and reduced head movement compared to uni-manual haptic and no-feedback conditions. In complex spatial search tasks, bi-manual haptics achieved performance on par with visual arrow cues while offering advantages in guiding users toward targets located behind them. Although participants reported slightly higher mental effort than with visual guidance, many indicated that the technique became more efficient with practice.

3.3.3. User Experience and Design Implications

The evaluation of alternative VR input methods, like haptic styluses, gloves, and embodied interfaces, reveals that user experience is shaped by the interplay of realism, control intuitiveness, and task context. Each study reviewed here contributes distinct insights into how interaction design can support or hinder immersion and usability in VR environments.

Bektaş et al. [5] demonstrated that embodied interfaces like the Limbic Chair offer benefits in reducing simulator sickness during flight simulation, but traditional gamepad controls still outperformed it in terms of usability, workload, and presence in city navigation tasks. This highlights a key trade-off: while embodied systems may enhance physical engagement or realism in some scenarios, they can introduce learning curves or performance penalties in others.

Palombo et al. [31] found that devices with high degrees of natural mapping, such as VR gloves, significantly increased users' sense of embodiment and cognitive absorption. Participants reported stronger spatial presence and more intuitive interaction with VR gloves than with standard controllers, urging that alignment of control inputs with natural motor expectations can deepen immersion.

Afzaal and Alim's [2] comparative study underlined how force-based haptics can improve precision and tactile perception in surface-based tasks, especially under occlusion. While handheld controllers allowed faster interaction in some tasks, the haptic stylus enabled smoother and more deliberate input, indicating a potential design strategy for balancing speed and accuracy depending on task demands.

Gibbs et al. [15] showed that the presence of haptic feedback, especially when combined with visual cues, enhanced users' spatial presence and perceptual awareness. Notably, haptic-only conditions sometimes outperformed visual-only ones in terms of user perception, further backing up the importance of multisensory integration in VR.

Taken together, these findings suggest that user experience in VR is not governed solely by realism or technological novelty but by the alignment of input modalities with the cognitive and perceptual models users bring to the environment. Designers should consider how natural mapping, feedback modality, and task complexity influence both performance and subjective experience. Embodied or haptic interfaces should be introduced with attention to learning curves and physical comfort to avoid compromising usability in pursuit of immersion.

In summary, haptic feedback offers powerful affordances for spatial learning, task engagement, and user immersion in VR. While more embodied interaction generally yields cognitive and experiential benefits, its effectiveness depends on careful integration with task demands, user familiarity, and ergonomic constraints. In the context of this research, SenseGlove-based interaction supports both pre-training and in-game engagement, aligning well with current best practices in embodied VR system design, albeit caution is warranted to ensure a smooth user experience.

4 Contributions

This chapter explicitly outlines the key contributions of this thesis, addressing notable gaps identified in existing literature regarding embodied interaction, spatial navigation, and haptic feedback within non-Euclidean virtual environments. This discussion begins with a comprehensive summary of these literature gaps, clearly motivating the research, and explicitly stating refined research questions and hypotheses. It concludes with an in-depth discussion of technical contributions involving the development and integration of innovative embodied and haptic interaction methods into the Holonomy VR environment.

4.1. Literature Gaps and Motivation

Despite significant research progress in VR-based spatial cognition, embodied learning, and haptic feedback technologies, notable gaps persist, particularly in the context of non-Euclidean (hyperbolic) VEs:

- **Limited Research on Embodied Pre-training in Non-Euclidean VR**
  Current literature extensively covers embodied interactions for enhancing spatial cognition within traditional, Euclidean VR contexts. However, minimal empirical evidence exists exploring structured pre-training interventions that leverage elements of embodiment explicitly tailored to spatial understanding and navigation efficiency within non-Euclidean, hyperbolic VR environments. It remains unclear how preliminary embodied interactions can facilitate the cognitive adaptation necessary to navigate non-intuitive, hyperbolic spaces effectively.
- **Comparative Evaluations of Embodied vs. Abstract Interfaces**
  Much of the existing literature comparing VR navigation interfaces has predominantly focused on fully embodied physical locomotion (e.g., walking) against simplified abstract interfaces (e.g., joystick or teleportation). Few studies examine nuanced variations within embodied control methods themselves, such as comparing drag-based embodied controls against abstract button-based interactions, particularly within challenging geometric scenarios such as hyperbolic environments.
- **Relationship Between Haptic Feedback, User Experience, and Spatial Learning**
  While prior studies separately address user experience, behavioural interactions, or objective spatial performance within VR, few explicitly correlate these data streams. The relationship between subjective user experience metrics (such as UEQ ratings), objective navigation performance (e.g., task completion times and accuracy), and behavioural data (interaction frequencies and durations) remains insufficiently explored, much less in contexts employing advanced haptic technologies.

4.2. Main Contributions

Building on the gaps identified above, this work aims to explore how embodied pre-training interventions and haptic feedback affect spatial understanding and usability in non-Euclidean virtual reality (VR).

While previous studies demonstrate that users can adapt to navigating hyperbolic spaces through exploration or repetition, it remains unclear how embodied interaction strategies and haptic affordances shape this adaptation process, particularly when users must perform in constrained, immersive, and counterintuitive virtual environments like Holonomy VR. This thesis makes two principal research contributions:

1. **Empirical Evaluation of Embodied Interaction in Pre-Training**
   This work introduces and evaluates a novel pre-training method using drag-based manipulation of a pawn in a grid-based minimap, leveraging embodied interaction through haptic gloves. The effectiveness of this intervention is compared to a more abstract button-based interaction. The goal is to examine whether embodied interaction facilitates better spatial cognition and adaptation to the non-intuitive nature of hyperbolic navigation.

2. **Assessment of Haptic Feedback in Spatial Navigation Tasks**
   By integrating the SenseGlove Nova 1, providing tactile and force feedback, into the navigation tasks in Holonomy VR, this thesis examines the extent to which haptic feedback influences user performance, cognitive map formation, and subjective experience in hyperbolic environments. This is measured through a combination of objective task metrics and the User Experience Questionnaire (UEQ).

4.2.1. Research Objective

The central objective of this thesis is to investigate how pre-navigation interaction methods and haptic feedback devices affect spatial learning and user experience in a non-Euclidean VR environment. Specifically, here, the haptic feedback device is leveraged to ground this interaction in embodied cognition. In particular, it seeks to assess whether embodied manipulation and tactile interaction can invoke familiar instincts while learning to facilitate the comprehension of such a complicated topic. By extension, it will be evaluated to what degree: the properties of hyperbolic geometry are understood, disorientation is mitigated, and cognitive mapping is improved.

For this investigation, participants engage in a structured training task followed by targeted navigation tasks within Holonomy VR that are built to utilise haptic feedback hardware. The training intervention is built to mimic the navigation task participants will be faced with in Holonomy VR, albeit in a low-fidelity, simplified environment focusing on the minimap and the navigation controls as seen in Figures 4.1 and 4.2. The reason for that is to isolate the navigation mechanics of Holonomy, to allow for familiarisation with the environment and the particularities of moving around in a hyperbolic world, with the ultimate goal of learning.

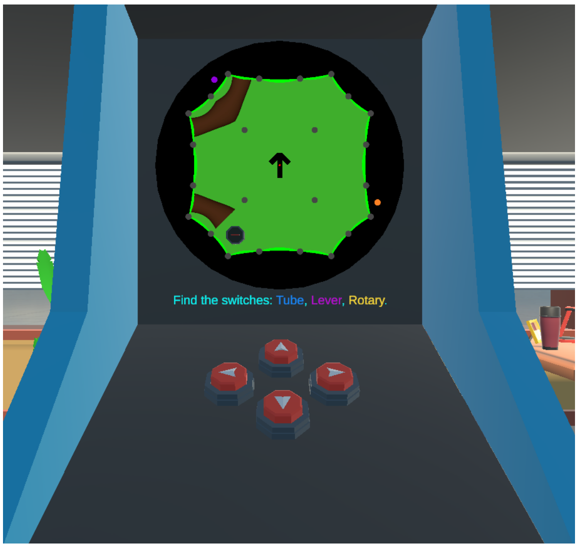


Figure 4.1: The training intervention environment using button-based controls [Control group].

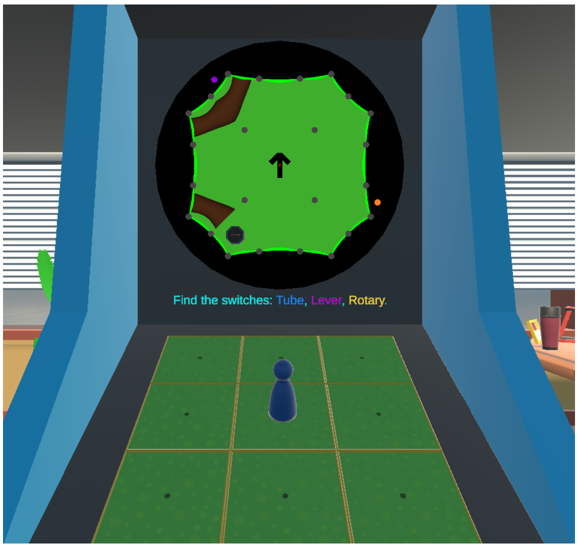


Figure 4.2: The training intervention environment using drag-based controls. [Treatment group]

The main takeaway for participants should be the method to reveal new tiles so that they can navigate to the switches both in the intervention and in the subsequent Holonomy VR tasks. Figure 4.3 shows how moving in a closed loop reveals new tiles on the minimap and rotates the world as a consequence of the holonomy property. Their performance will be measured on the basis of completion times and efficiency of navigation, which is defined as the least amount of deviation from the optimal path to each of the switches to be collected. The performance scoring is explained in a lot more detail in Section 6.2. Exposing participants to the properties of the world in this manner would allow them to carry over this gathered knowledge to the actual hyperbolic environment, and a discrepancy in performance between the two training groups (buttons and dragging) would be observed, should the embodied approach have any effect on the training.

(a) Origin **(b) Moving Up ↑** **(c) Moving Right →** **(d) Moving Down ↓** **(e) Moving Left ←**

Figure 4.3: Moving in a closed loop (↑, →, ↓, ←) starting at the origin to reveal new tiles. The world is also rotated 90° and the user does not arrive back at the origin.

4.2.2. Research Questions

The thesis is guided by the following research questions, each designed to address a specific dimension of the navigation problem in hyperbolic VR:

RQ1: *How does embodied learning (drag-based interaction with haptic gloves) affect spatial understanding and navigation performance in hyperbolic VR compared to abstract button-based interaction?*
This question is motivated by the strong evidence that active, body-involving exploration enhances spatial cognition [26][38], and addresses the cognitive and behavioural impact of embodied learning principles in this context.

RQ2: *How does variation in user-driven interaction intensity with environmental elements in a haptically-enabled, non-Euclidean VR environment relate to user experience, engagement, and usability perceptions?*
This question builds upon interface research demonstrating that active, multimodal interaction, particularly through tactile feedback, is associated with enhanced user satisfaction and presence [31][15]. While this study does not experimentally manipulate the presence or absence of haptic feedback itself, it explores how naturally arising variations in participant-driven interaction intensity, enabled by the availability of tactile feedback, relate to experiential outcomes such as immersion, engagement, and perceived usability. Such an approach acknowledges individual differences in exploratory behaviour, providing insights into whether more active physical exploration, through increased tactile interactions with reactive environmental elements, correlates positively with subjective user experience and potentially mitigates spatial discomfort in non-Euclidean virtual environments.

4.2.3. Hypotheses

To address these research questions, the following hypotheses are proposed:

Hypothesis 1 (H1):
Embodied-trained participants will exhibit superior navigation performance, e.g., faster completion times, and more optimal path choices, compared to non-embodied trained participants using button-based interactions.

This hypothesis is grounded in the theory of embodied cognition, which emphasises that learning and memory are enhanced through sensorimotor engagement with a task environment [50][26]. Prior studies show that embodied navigation (e.g., full-body motion, gesture-based input) improves spatial memory and pathfinding skills in VR settings [38][8]. While such evidence exists for Euclidean contexts, the application to hyperbolic space is novel. This thesis bridges this gap by testing whether similar benefits arise in non-Euclidean spaces where spatial logic is unfamiliar and cognitive maps are harder to build [48].

It is expected, then, that embodied-trained participants will require fewer moves to reach targets, make fewer navigation errors, and report a more intuitive understanding of movement mechanics and hyperbolic properties. The feedback of each participant will be quantified through the UEQ responses [17], and scored questions about their takeaways regarding the hyperbolic world.

Hypothesis 2 (H2):
Participants who voluntarily spend more time interacting with reactive environmental elements (e.g., touching trees and bushes) are expected to report higher levels of perceived engagement, immersion, and overall positive user experience compared to participants who engage minimally or not at all. Additionally, perceived interaction intensity is hypothesised to correlate positively (though possibly modestly) with actual measured interaction time and will also be associated with higher User Experience Questionnaire (UEQ) scores, particularly on scales such as efficiency, perspicuity, and stimulation.

While all participants used identical haptic glove hardware, this study specifically explores how naturally arising variation in user-driven interaction intensity may inform our understanding of active sensorimotor engagement's role in shaping subjective experiences in non-Euclidean virtual reality. Drawing on embodied learning theory, it is theorised that participants' physical exploration and tactile interaction with reactive elements could foster stronger sensorimotor coupling and enhanced affordance perception [1][8]. This notion aligns closely with the concept of ”learning by doing” or ”enactive exploration,” which posits that sensorimotor activity is not merely functional but fundamentally instrumental to cognition and affective experiences [50].

Prior research further supports this hypothesis by indicating that tactile feedback from non-essential virtual objects can positively influence users' sense of engagement and environmental richness, even when such interactions are not strictly required for task completion [40]. This study investigates these ideas through two primary data streams:

- Objective interaction time with reactive objects (bushes, trees).
- Self-reported perception of interaction intensity and environmental richness.

By correlating these measures with responses from the UEQ (particularly efficiency, perspicuity, and stimulation scales) and additional custom items addressing memorability and presence, this study aims to provide nuanced insights into how voluntary physical exploration, beyond navigational necessity, may meaningfully shape user experience in non-Euclidean VR.

4.3. Technical Contributions

In preparation for answering the aforementioned research questions, the Holonomy VR project requires some modifications and updates. Primarily to support and facilitate the use of the SenseGlove and enable haptic-focused interactions with the environment.

4.3.1. Minimap Augmentation

The minimap in Holonomy VR is the main tool for navigating the hyperbolic space; it gives information as to where the landmarks are, and it updates as the player moves along. To fit in with the theme of the intervention and to make it more visually appealing, the minimap was updated to support the use of 3D models for landmarks.

The now updated minimap uses the Poincaré tile centres, and the Möbius transformation to the centre of the minimap to map 3D models representing the switches, and shift them accordingly as the player moves. The achieved effect is this 2.5D version of the otherwise flat minimap, as it is illustrated in Figure 4.4. To improve performance, only landmarks in the immediately visible tiles within the boundaries appear as three-dimensional.

The indicators of the landmarks now support colour-coding to make it easier to follow their movements when they are beyond the boundaries of the 9-tile grid. Previously, all landmarks appeared as red beacons, which have now been updated to support any colour for any landmark. The user interface is also simplified, stating only the remaining landmarks to be collected, and displays them in their respective colours that match their beacons. This creates a much less confusing setup for the player, and it is illustrated in Figure 4.5.

Figure 4.4: The 3D models of the switches plotted on the minimap to achieve a 2.5D effect.

Figure 4.5: The colour-coded beacons of the switches outside the bounds. The remaining objectives text also matches the switch-specific colours.

4.3.2. Tutorial

To make the provision of instructions to participants consistent across trials, an automated tutorial scene was created. Participants are placed in a scene in front of a whiteboard, where they receive instructions regarding the intervention phase and its goal, what they are expected to do, and how they navigate depending on the group they were assigned to.

A sequence of instructions was choreographed and animated; first, the goal of the intervention is explained, and the participants are introduced to the switches that act as their objectives. They are then shown what they look like on the minimap (see Figure 4.6), followed by looping animations of the two different control schemes (see Figure 4.7). The tutorial then urges them to ask any remaining questions before prompting them to press a button to finish this segment and move to the next phase. The entire sequence is narrated to make it easier for the participants to follow along.

Figure 4.6: Tutorial segment with animated minimap annotations explaining what the switches look like.

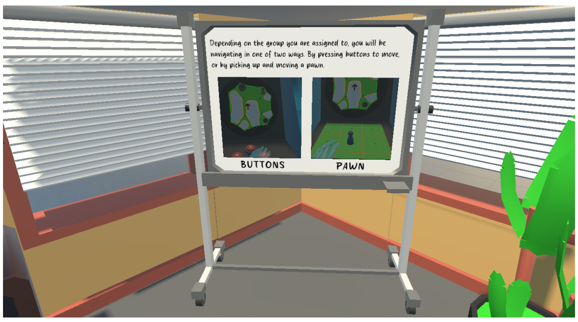


Figure 4.7: Tutorial segment with looping animations explaining the control schemes for the intervention phase.

4.3.3. Intervention

The intervention phase of the study is a completely new addition to the project, made for the purpose of this study. It utilises core components of Holonomy VR, such as the minimap, while adding new functionality aimed at facilitating learning about moving in the hyperbolic environment.

Surrounding environment

The intervention scene was custom modelled to represent a scene from an office, as shown in Figure 4.8. This was made in an attempt to make the surroundings more welcoming. The participants interact with an arcade machine that is placed in the middle of everything, with the minimap acting as a screen, and the controls right below it.

Figure 4.8: The modelled environment of the intervention scene.

Button-based navigation

One of the control methods for the participants in the training intervention is pushing buttons using their virtual hands, utilising the haptic functionalities of the SenseGlove. The buttons were created from scratch and were made to mimic the behaviour of a spring-loaded button as closely as possible (see Figure 4.9). The buttons, when activated, emit a small vibration to the hand that initiated the activation through the motors that the SenseGlove is equipped with. They are also fully force-feedback enabled, meaning that the hand feels resistance when pushing against them.

Four of these buttons, arranged in a cross pattern, make up the controls for one of the intervention groups (see Figure 4.10). A button exists for each of the four cardinal directions: up, down, left, and right, and pushing each respective one moves the player one tile in that direction.

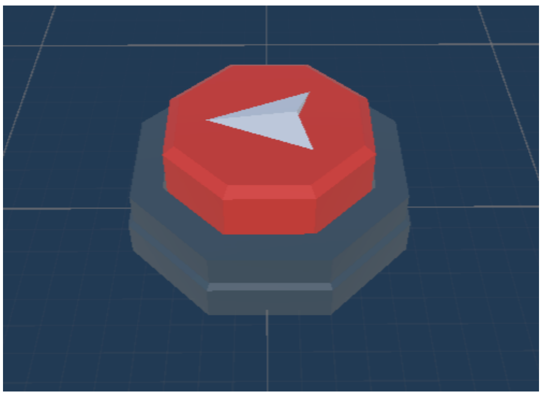


Figure 4.9: Button that is used for controlling the player in the intervention phase.

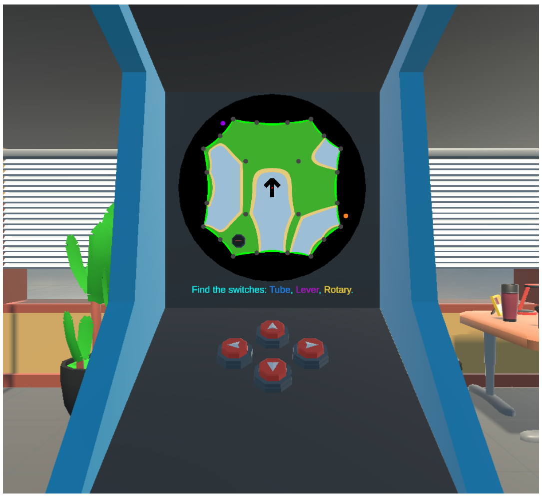


Figure 4.10: Setup of the button-based control group during the intervention phase. The minimap is located at the top, with the 4 buttons right below it.

Drag-based navigation

The second control method for the participants during the intervention phase is picking up a pawn piece and dragging it across a 9-tile grid board using their virtual hands. This method is rooted in embodied learning, and the movements their hands and arms execute are meant to make understanding the hyperbolic environment navigation easier and more memorable.

The board and pawn, as seen in Figure 4.11, were made to utilise the SenseGlove features fully. The 9-tile grid corresponds exactly to the tiles the players will be placed on in the main game of Holonomy VR. The red rectangle around the board is created using a custom shader to simulate a hologram barrier; it is hidden during runtime and only appears when the pawn is taken out of bounds. When the pawn is out of bounds, it is dropped by the hand currently grabbing it, and it is snapped to the centre of the closest tile. This is to prevent illegal moves while the pawn is off the board.

The pawn offers full force-feedback support, and the hand that grabs it snaps to pre-defined points around it to give a natural impression of picking up a board game pawn. The participants move around the hyperbolic environment by dragging the pawn around; moving the pawn across tiles moves and updates the minimap. On every tile cross, there is a very small vibration emitted through the fingertips of the glove grabbing the pawn.

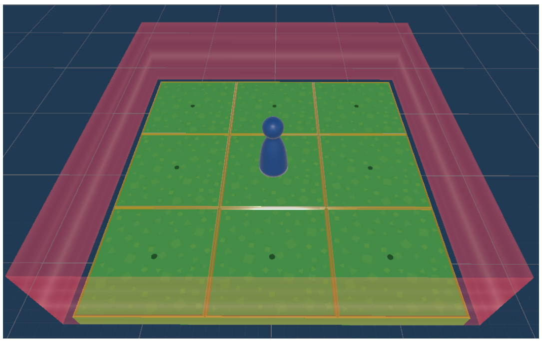


Figure 4.11: Board and pawn used for controlling the player in the intervention phase. The holographic barrier here is enabled.

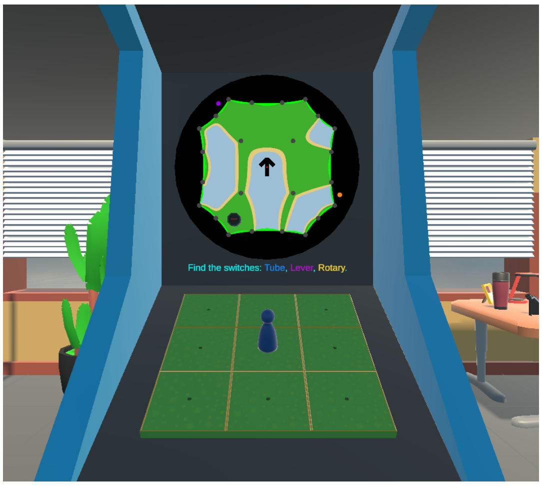


Figure 4.12: Setup of the drag-based control group during the intervention phase. The minimap is located at the top, and the board with the pawn right below it.

4.3.4. Haptic Switches

Pivotal additions to the Holonomy VR project are the haptic switches that were created to replace the landmarks that previously acted as objectives. The goal was to create objectives for the main game that forced some form of interaction with the SenseGlove. Hence, the following switches were created:

- **Lever switch:** Pull laterally until the switch reaches the activation point.
- **Rotary switch:** Twist clockwise or counter-clockwise until the switch reaches the activation point.
- **Tube switch:** Pull vertically until the switch reaches the activation point.

Each of the switches, as can be seen in Figures 4.13, 4.14, and 4.15, has a red handle where the SenseGlove can grab on to initiate the activation movement. All switches are force-feedback enabled; reaching the activation point for any of them will lock the switch in place and will, in turn, send a vibration through the hand that was grabbing the switch. If any of the switches are released before reaching their activation point, they will spring back to their initial position. Each of the switches was placed on a pedestal to make it easier to interact with while in the Holonomy VR environment, since this positions them around waist-height.

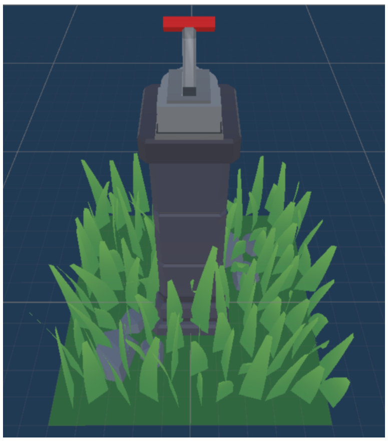


Figure 4.13: Lever switch tile.

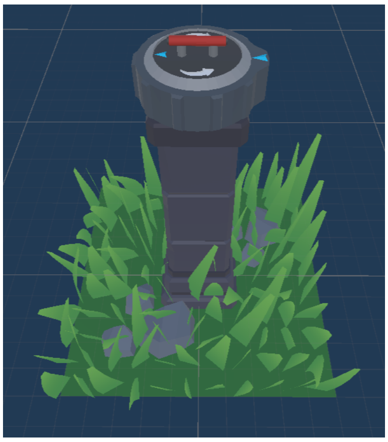


Figure 4.14: Rotary switch tile.

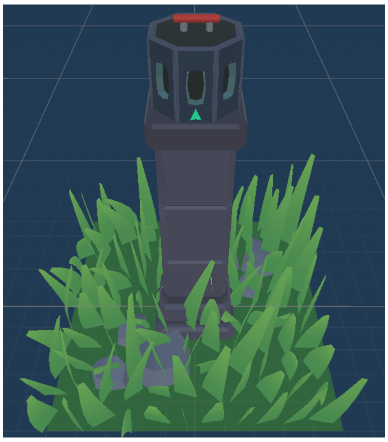


Figure 4.15: Tube switch tile.

4.3.5. Haptic-enabled Environment

To make the environment in Holonomy VR feel more alive, the trees and the surrounding hedge in the playable grid area offer vibrotactile feedback. The tree trunks give off a small impact vibration when they are touched or grabbed, as well as full force-feedback resistance. Similarly, the surrounding hedge causes the hand that collides with its bounding box (see Figure 4.16) to vibrate while the two colliders overlap. The vibration scales in intensity depending on the depth that the hand penetrates the bounding box and is calculated with $Intensity = A$ (see Equation 4.1):

$$\tilde{F} = F_{\text{front}} - s\,\Delta d, \tag{4.1a}$$
$$r = s\left(x - \tilde{F}\right), \tag{4.1b}$$
$$d = \max\!\left(0, \min\!\left(\tfrac{r}{t}, 1\right)\right), \tag{4.1c}$$
$$A = k\,d, \tag{4.1d}$$

where

$$F_{\text{front}} : \text{frontFace}, \quad \Delta d : \text{depthOffset}, \quad s \in \{+1, -1\} : \text{dirSign},$$
$$x : \text{axisValue}, \quad t : \text{thickness}, \quad k = 0.65\ \text{vibrationGain}.$$

These interactions are not mandatory to complete the levels in Holonomy VR but serve a secondary role in enriching the experience for the player. The environment being reactive to the player's presence can help with feelings of claustrophobia and can increase engagement with the activity at hand.

Figure 4.16: The bounding boxes on the surrounding hedge and trees in the Holonomy VR environment.

4.3.6. Environment Interaction Tracking

To answer the second research question, it is necessary to collect data regarding the players' interaction with the environment. To that extent, another system was created that produces another data stream with interaction data of the player with the trees and hedges. For each level in the study, a log is maintained that tracks the total interaction time of the player with each of the two haptic-enabled elements of the environment (excluding the switches). Interaction here is defined as any of the colliders on the hands intersecting with any of the colliders of the trees or hedges. The total interaction time for the trees and hedges is stored separately and is updated every frame, providing millisecond (1 ms) accuracy.

4.3.7. Experiment Manager

For administrative purposes, and to connect all the different scenes of the experiment together, an experiment manager was developed that takes care of data logging and transitioning between scenes. Data is logged automatically depending on which phase of the study is completed. The data includes a participant ID, the phase they completed, time taken, the complete path taken, steps away from the hyperbolic origin, interaction time with the environment, objectives and the order they were collected in, and a timestamp. The data is stored in an Excel file for further analysis. Through the experiment manager, the training group can be selected, which automatically changes all the necessary settings in every scene to accommodate it. A participant ID can also be entered before the experiment commences, which should match the participant ID entered in Qualtrics, where the survey takes place.

The experiment manager starts the experience in the tutorial scene and then awaits the prompt from the participant to move on to the next phase. It then moves to the first part of the intervention phase. Once complete, it presents the participant with an end screen and a prompt to call the researcher to move on to the next step of the intervention. Once the intervention is complete, another screen is shown so that the participant can remove the hardware and complete the first UEQ. After equipping the hardware again, the participant goes on to complete 2 levels in the Holonomy VR environment, which, with the completion of the final survey, marks the end of the study. More information regarding the exact experiment procedure is outlined in the next Chapter 5.

5
Methodology

This chapter explains how the study was designed and executed. The between-subjects experiment is outlined, which compares Buttons (conventional) vs Drag (embodied) training with random assignment in a hyperbolic VR setting. This is then followed by a matched Main evaluation in Holonomy VR. The participants and demographics are then described, along with the materials, apparatus, and telemetry captured. The procedure is detailed across phases: pre-Intervention, Intervention, and Main. The chapter concludes with a brief mention of the pilot and logistics setup to ensure consistency and reproducibility.

5.1. Experimental Design

This study employs a between-subjects experimental design with a single independent variable at two levels: the training method (Embodied vs. Conventional). Participants are randomly assigned to one of two intervention conditions to evaluate the impact of embodied interaction on spatial navigation performance and user experience in a hyperbolic VR environment. The details of this intervention, as well as subsequent main tasks in the Holonomy environment, are thoroughly described in the following sections.

5.2. Participants

Participants are recruited from a university setting, primarily targeting students with diverse backgrounds in education, experience with VR and video games, and varying familiarity with hyperbolic geometry. Demographic information such as age, gender, educational background, field of study (based on ISCED-F[1]), and prior experiences are collected via a survey to ensure balanced distribution between experimental conditions. This information is necessary to form relations between participant performance and their distinct characteristics that could be affecting it. For example, someone very experienced with VR or video games exhibits a significant advantage over other participants simply because they don't need that much acclimation time.

5.3. Materials and Apparatus

The experiment utilises the Meta Quest 2 VR headset paired with SenseGlove Nova 1 haptic gloves. All participants use the gloves throughout the entire experiment. The experimental software consists of a custom-designed intervention using a simulated arcade machine interface for the learning phase and the Holonomy VR environment for the main navigational tasks. Since Holonomy VR was a project made with Unity, all additions were created in the same environment using Unity version 2022.3.27f1. The majority of the 3D assets were custom-made using Blender v4.3; assets for the intervention environment were adapted from an office pack in the Unity AssetStore[2], as well as some additional low-poly

[1]https://ec.europa.eu/eurostat/statistics-explained/index.php?title=International_Standard_Classification_of_Education_(ISCED)
[2]https://assetstore.unity.com/packages/3d/characters/low-poly-office-pack-characters-props-119386?aid=1100liZev

assets adapted from Poly Pizza[3]. Both systems log detailed data about user interactions, task performance, paths taken, and times spent interacting with various environmental elements.

5.4. Procedure

The experimental procedure consists of three distinct phases: pre-intervention, intervention, and post-intervention, as seen in Figure 5.1 below.

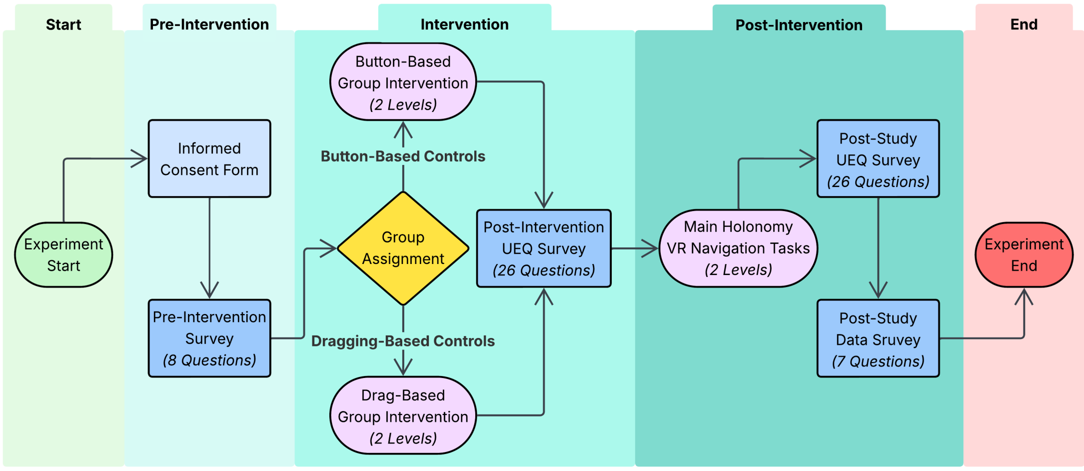


Figure 5.1: The procedure followed for the experiment. Ellipses represent actions by the participants involving the use of the VR headset and haptic hardware. Rectangles represent forms and surveys filled in by the participants. The rhombus represents the group assignment point, where participants are divided into groups.

5.4.1. Pre-Intervention Phase

Upon arrival, participants provide informed consent via a form. The form outlines the procedure for the experiment and briefs participants about the collected data, participation risks, and voluntary participation rights, followed by a chance to ask any questions that may arise. The full informed consent form can be found in Appendix C. Afterwards, they complete an initial survey gathering demographic data, prior VR and gaming experience, familiarity with hyperbolic geometry, and self-reported claustrophobic predispositions. This part of the survey can be found in Appendix B.1.

5.4.2. Intervention Phase

Participants are then randomly assigned to either:

- **Conventional (Control) group**: Participants interact using abstract button presses corresponding to cardinal movements of the player in the hyperbolic world.
- **Embodied (Treatment) group**: Participants interact by physically dragging a virtual pawn across a 3×3 grid that corresponds to the grid in the main Holonomy VR game.

During the intervention phase, participants observe the hyperbolic environment's reactions via a minimap, relating their movement actions to changes within the non-Euclidean space. The tasks for both groups are identical; they have to reach 3 objectives across 2 levels of increasing difficulty. Difficulty here is defined as the furthest from the hyperbolic origin that an objective exists in the level. For the first level in the intervention, the objectives lie at most 3 tiles away from the hyperbolic origin. In the second level, that limit is increased to 5 tiles from the hyperbolic origin. This is to allow for gradual familiarisation with both the control mechanisms of the system and with the hardware itself. Since completion times may vary across participants, a cut-off point has been set at 15 minutes for each level of the intervention phase to prevent the session from running over the estimated duration of 60 to 70 minutes. These data entries will be classified as partial or invalid, but could still hold some value in giving insight regarding the second research question about user experience.

[3] https://poly.pizza/

Following the completion of the two levels in the intervention, participants are asked to fill in the first part of the survey, as seen in Appendix B.2. This consists of 26 questions that make up the user experience questionnaire (UEQ) [17], which is a widely used tool for measuring user experience with interactive products. It assesses user perceptions across six key dimensions: Attractiveness, Efficiency, Perspicuity, Dependability, Stimulation, and Novelty. This is a concise and efficient way to gather quantitative data regarding the users' experience and will be used twice in this study. The participants are asked to fill in only this part of the UEQ according to the experience they've just had during the intervention phase, in isolation, with no other considerations about the study. The survey at this point also serves as a nice break for the participants to keep them from exhausting themselves due to the lengthy nature of the study and the extra weight of the hardware.

5.4.3. Post-Intervention Phase (Holonomy VR)

The next step after the intervention is the post-intervention phase, where participants engage in two navigational tasks within Holonomy VR. Once again, the participants are required to complete two levels of increasing difficulty, both of which have 3 objectives/switches to be found. Difficulty here is defined in the same manner as the intervention phase, namely, how far from the hyperbolic origin the switches exist. For the post-intervention (Main) levels, these limits were set to 4 tiles away from the hyperbolic origin for level 1, and to 6 tiles away for level 2. These limits were chosen as a step-up from the intervention phase since the participants are expected to carry over their gathered knowledge and experience. A cut-off point was also chosen for this part of the study to keep participants from exceeding the estimated total duration. This is set at 15 minutes for each of the levels, similar to the intervention phase.

All participants continue using the SenseGlove Nova 1 as they have been for the entirety of this study. The environment in Holonomy VR is haptic-enabled, as explained in Subsection 4.3.5, meaning that it reacts to the participants' presence. The trees and bushes on all levels react to the player's touch, providing a more engaging and memorable experience. This functionality will be disclosed to all participants to make them aware it exists, since the second research question requires data about their interaction with these elements of the environment. That behaviour is not forced, though, and all participants will be left to their own devices to act as naturally and as instinctively as possible when tackling the navigation tasks of this segment. This interaction is, in addition, not mandatory to complete any of the levels; the only haptic interaction required is with the switches that are encountered throughout the levels. Participants need to seek the switches located around the hyperbolic world, and once found, they are required to activate them (in the ways described in Subsection 4.3.3) to collect them and move on to the next objective.

When the two levels in Holonomy VR are completed (or the cut-off is reached), participants are again required to fill in the second part of the survey. This is 26 additional questions comprising a second UEQ, whereas this time it is targeted at the post-intervention only. Participants are asked to focus solely on the Holonomy VR experience and answer according to that, without regarding any previous parts of their session, so that the Holonomy VR part of the study can be analysed in isolation.

The final part of the experiment is the remaining questions in the survey, which are designed to assess different areas of the participants' understanding regarding the hyperbolic world and its properties. The questions, as listed in the survey, can be found in Appendix B.3. The full version of the survey can be found in the Appendix B. Specifically, the last part of the survey is structured as follows:

- **Hyperbolic World Understanding (4 pts)**
  (4) Multiple-choice questions targeted to assess the participants' understanding of the hyperbolic world and its properties.
- **Self-reported Environment Interaction (5-point scale)**
  (2) Questions made for the participants to report how they perceived their interaction in the Holonomy VR environment.

- **Self-reported Feelings of Claustrophobia (5-point scale)** (5) Questions adapted from the claustrophobia questionnaire (CLQ) [36] (available online[4]), aimed at assessing any feelings of claustrophobia that participants may have experienced during the study

This brings up the total to 11 questions to round off the last part of the session. With this, the data collection is complete and the session is concluded.

5.5. Pilot User Study

To verify the feasibility of the study and to test the robustness of the system, a very small pilot study with 4 participants was conducted. The duration of the session was left open-ended to gauge an estimate of the total duration of the final form of the study, but would not exceed 90 minutes. The participants were students and professionals, aged between 24 and 26 years old, 2 males and 2 females. The study was conducted on private premises, with the system and hardware that were used to develop the entire system, to minimise any possibilities of unintended behaviour being introduced due to changes in the equipment. The participants were split into groups of two, with group one training using the button-based controls and group two training using the drag-based controls. The pilot study served as a limit test to determine the difficulty of the levels for both the intervention and post-intervention phases, as well as hone in on a suitable cut-off time for any prolonged attempts, which 15 minutes was found to be sufficient.

A lot of feedback from the participants in terms of usability and interface design was incorporated to accommodate some poorly designed features. For example, the timings in the tutorial segment were found to be too short to follow for someone experiencing it for the first time. Additionally, the placement of the minimap in the Holonomy VR segment was adjusted to make viewing the environment slightly easier, since the minimap follows head movements and is constantly in view. The pawn in the drag-based controls was found to be too difficult and unnatural to grab, so its snap points were adjusted. Furthermore, a lot of unintended game-breaking behaviour was discovered and resolved. A participant managed to activate a switch while not being in its tile; after crossing between tile boundaries and while still holding on to the switch, the participant's hand got unparented from the XR Rig, causing it to float infinitely in the level. This was due to the Holonomy implementation deleting and re-spawning tiles while the player is crossing tile boundaries, and since the hand was parented to one of the tiles, it got removed with it. A solution for this issue was found by temporarily disabling the grabbing function for any hands currently holding on to a switch when the player is crossing tiles. The data logging was also faulty when participants did not manage to find all objectives on time; thus, some fail-safes were set in place that can restore the session data through some researcher-controlled shortcuts.

All participants were able to complete all levels in under 13 minutes, albeit with some variations in level difficulty, but this serves as a good indicator that the cut-off point is well picked. The survey questions were also answered in a reasonable amount of time by everyone, meaning that the overall session time shouldn't exceed 70 minutes. The different strategies of each participant proved to be a good exercise for determining when a hint should be given that they are heading too far away from their objective. Since no precise tile distances are shown, it is not possible to determine an exact moment to give a hint for a change in course, but a visual indication when the switch appears to be moving much further was found to be a good moment for it.

5.6. User Study

The user study was conducted at TU Delft facilities. Some of the sessions took place at the XR zone within the TU Delft library, while most sessions took place at the Insyght lab within the EEMCS faculty. The setup at both locations was identical: a 3×3 meter area with the centre of it marked to accurately place participants across phases. A monitor was also made available for the responsible researcher to see a live feed of the participants' point of view, in order to answer any questions or provide additional guidance.

[4] https://psytests.org/diag/clqen.html

Consistency across sessions was ensured through the scripted tutorial phase in Unity, as mentioned in Subsection 4.3.2. Only one responsible researcher conducted all the sessions, so no coordination was required. Besides the tutorial phase, the same information was given to all participants. The manoeuvre of performing a loop around a vertex to ”unlock more tiles” was given as an initial hint to everyone. All sessions followed the same structure as outlined before in 5.4, and the data from all participants are presented in the next chapter.

6

Results

This chapter reports empirical findings in the order needed to support the two research questions. It begins with sample characteristics and balance checks for demographic confounds. The performance scoring framework is then defined with three robust within-level pillars that are used to obtain a composite metric. Next, performance is compared across groups and phases using distributional graphics and pillar profiles to pinpoint where differences arise. The focus then narrows to the Main phase, where groups are on equal footing. Parametric and rank tests, a mixed-effects model, and two sensitivity analyses are presented. This is then followed by exploratory checks to contextualise robustness and investigate parallel avenues. Finally, for the second research question, the interaction intensity and user experience in the Main phase are examined. A validity link between logs and self-reports is then established. To track perceived engagement, the UEQ scales and item-level regressions are tested. Throughout, estimates are shown with uncertainty, and multiplicity is controlled where families of tests are run.

6.1. Demographics

A total of 28 participants took part in the user study. Participants were divided evenly across the two groups; 14 were assigned to the drag controls group, and 14 were assigned to the buttons control group. Each group had 11 male participants and 3 female participants. Demographic balance was checked descriptively and with a non-parametric Mann-Whitney U test per variable. A smaller U value indicates a greater difference between distributions, with the associated p-values indicating statistical significance. No large imbalances were observed that would warrant covariate adjustment specific to demographics, since no p-value falls under the significance level $\alpha = 0.05$. Table 6.1 shows all the checks, with just the Age and Hyperbolic geometry experience variables being close to statistical significance. The distributions for the "Field of study" and "Educational attainment" can be seen in Figures 6.1 and 6.2, respectively. The Education distributions were almost identical, and the field of study distributions are also comparable. Important to note is that no participant exceeded the 15-minute cut-off limit, as mentioned in Section 5.4. This means that the full range of obtained data is used for the following analysis.

Demographic info	U statistic	p-value
Age	65.000	0.125
Gender	98.000	1.000
Field of study	115.000	0.430
Educational attainment	104.000	0.783
Video game experience	96.000	0.943
VR experience	88.500	0.652
Hyperbolic geometry experience	120.500	0.286
Claustrophobia predisposition	105.000	0.724

Table 6.1: Mann–Whitney U test for demographic balance checks across groups.

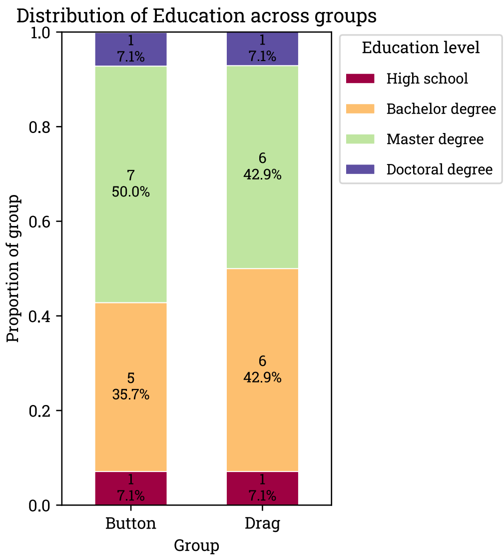


Figure 6.1: Distribution of educational attainment across the two groups.

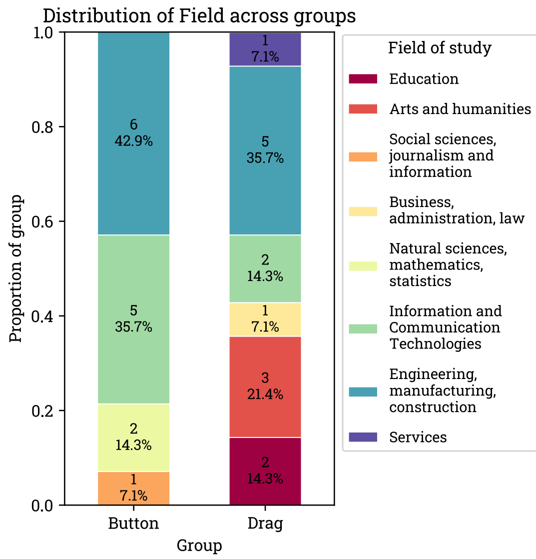


Figure 6.2: Distribution of field of study across the two groups.

6.2. Performance Scoring

A robust metric was required to measure participant performance. A composite metric was developed that combined three pillars of scoring, each targeting a different facet of a participant's run through a level.

- **Speed:** How fast a participant completed a level
- **Efficiency:** How long (in tile counts) was a participant's path compared to an optimal path*
- **Sequence:** How did a participant's path sequence of tiles compare to an optimal path*

**The notion of the "optimal path" is explained in the next subsection*

6.2.1. Optimal paths

To obtain an optimal path to compare the participant's paths to, an A* algorithm was used to obtain approximations of the optimal path from point to point. There are 3 objectives in every level, in every phase, and all levels start from the hyperbolic origin of the world. This means there are 6 permutations of the objective order collection for every level (e.g O → A → B → C, O → B → C → A, ...).

A caveat with this approach is the environment that A* considered within Holonomy VR. The path approximations were obtained from an unrestricted version of the tile-based world of Holonomy VR; that is, the shortest path from point to point without accounting for tile revealing or rotations to unlock more tiles, as a participant would have to do. This outputs a theoretical approximation of the shortest paths to collect all objectives, which is not necessarily reproducible in live Holonomy conditions.

To circumvent this shortcoming, when picking which permutation to consider for each participant, the permutation that maximised the sum of the efficiency and sequence pillars was chosen. This is to avoid over-penalising viable strategies and to use the best possible optimal path approximation as a starting point for the pillar calculations. This is applied identically to both groups. The calculations of each pillar are explained in the next subsections.

6.2.2. Speed pillar

The speed pillar reflects how quickly a participant completed a level. Each participant's time was converted to seconds, and a leave-one-out median $LOO_{median_{i\ell}}$ was computed per participant $t_{i\ell}$ time for the same level ℓ (see Equation 6.1a). Levels differ in inherent difficulty and length, so raw times are not comparable across levels. The $LOO_{median_{i\ell}}$ is chosen because this value will be used to calculate the deviation from the ”typical” time for that level. Excluding the currently considered observation from the median calculation does not artificially shrink this deviation. That is, participants essentially don't set their own benchmark. The median is used since it is robust to outliers. A log relative time $LRT_{i\ell}$ was formed using the deviation of the participant's time $t_{i\ell}$ from the typical $LOO_{median_{i\ell}}$ time for this level. This gives the Equation 6.1b. $-LRT_{i\ell}$ is used so higher means better; $\log()$ gives symmetry for multiplicative differences and makes the effect additive (see Equation 6.1c).

$$\mathrm{LOO}_{\mathrm{median}_{i\ell}} = \mathrm{median}(\{t_{j\ell} : j \neq i\}) \tag{6.1a}$$

$$\mathrm{LRT}_{i\ell} = \log(t_{i\ell}) - \log(LOO_{median_{i\ell}}) = \log(\frac{t_{i\ell}}{LOO_{median_{i\ell}}}) \tag{6.1b}$$

$$-\mathrm{LRT}_{i\ell} = \log(\frac{LOO_{median_{i\ell}}}{t_{i\ell}}) \tag{6.1c}$$

6.2.3. Efficiency pillar

The efficiency pillar compares how close the participant's path length L_{real} was to the best optimal path approximation length L_p^*. As explained in 6.2.1, the considered optimal path is the permutation that maximises both the efficiency and sequence components. Here, inefficiency F_p for participant p is used instead (see Equation 6.2b), so that 0 is considered optimal. An inefficiency score of 0.25 would mean 25% longer than the optimal path length and can be interpreted directly. The sign is flipped again $-F_p$ so that higher means better.

$$\mathrm{R} = \frac{L_{real}}{L^*} \tag{6.2a}$$

$$\mathrm{F_p} = R - 1 = \frac{L_{real} - L_p^*}{L_p^*} \tag{6.2b}$$

6.2.4. Sequence pillar

The sequence pillar reflects how much the participant's path tile sequence deviated from the best optimal path approximation tile sequence. Similarly to the efficiency pillar, the considered optimal path is the permutation that gives the best combined sequence and efficiency score. The sequence score is a composite of two equally weighted components:

- J: Jaccard bigram similarity. It captures local order, immediate transitions by the participant, but does not look at far-apart sections of the sequence. It penalises reversals, detours, and extra steps. J is calculated as the ratio between the sum of the $\min$ count of bigram g occurrences in the participant's path $\mathrm{count_p}(g)$ and optimal path $\mathrm{count_{opt}}(g)$, over the sum of $\max$ (see Equation 6.3a).
- $\mathrm{LCS_{F1}}$: F1 score of the longest common subsequence. It captures global order and rewards equal relative order, even if extra tokens (tiles) are inserted or some are skipped. It penalises long-range inversions with truly scrambled order. $\mathrm{LCS_{F1}}$ is calculated from the precision P and recall R scores (see Equation 6.3b). The terms combine the longest common subsequence LCS with the participant's path length L_{real} and optimal path length L^* (see Equation 6.3c).

The two metrics complement each other, which is why they are used together to obtain a final $\mathrm{SEQ_p}$ score (see Equation 6.3d). The metrics are conceptually similar; if they correlate moderately-strongly, and positively, they are plausibly measuring the same construct. In that case, an equal weight of 0.5 is justified. Table 6.2 shows a strong positive correlation since $\alpha > 0.9$ and Pearson's $r > 0.8$ across Intervention, Main, and pooled (all runs) phases.

$$J = \frac{\sum_{g \in G} \min\{\text{count}_\text{p}(g), \text{count}_\text{opt}(g)\}}{\sum_{g \in G} \max\{\text{count}_\text{p}(g), \text{count}_\text{opt}(g)\}} \in [0,1] \tag{6.3a}$$

$$\text{P} = \frac{LCS}{L_{real}}\ , \ \text{R} = \frac{LCS}{L^*} \tag{6.3b}$$

$$\text{LCS}_\text{F1} = \frac{2 \cdot P \cdot R}{P + R} \tag{6.3c}$$

$$\text{SEQ}_\text{p} = 0.5 \cdot LCS_{F1}\ +\ 0.5 \cdot J\ , \quad \text{SEQ}_\text{p} \in [0,1] \tag{6.3d}$$

Phase	Cronbach's α	Pearson's r
Intervention	0.909	0.833
Main	0.958	0.920
Pooled	0.945	0.896

Table 6.2: Cronbach's α and Pearson's r for J and LCS_F1 across phases.

6.2.5. Standardisation

All 3 of the aforementioned pillar scores are then standardised using a robust z-score. This is necessary to obtain a performance composite, since all pillars start on different scales; log-ratio for $-LRT_{i\ell}$, per cent overhead for $-F_p$, and $[0,1]$ for SEQ_p. Standardisation happens within-level to remove inherent level difficulty/length effects. The mean $\bar{x}$ and SD s are sensitive to outliers; with an $N = 28$, only two runs per level, and mostly heavy-tailed data, a couple of extremes can dominate $\bar{x}$ and s and can inflate a classic z_i score (see Equation 6.4). Instead, the $\text{median}(x)$ and mean absolute deviation MAD are used to obtain a robust z-score z_i^{rob} (see Equation 6.5).

$$z_i = \frac{x_i - \bar{x}}{s} \tag{6.4}$$

$$\begin{aligned} \text{MAD} &= \text{median}\big(|x_i - \text{median}(x)|\big) \\ z_i^\text{rob} &= \frac{x_i - \text{median}(x)}{1.4826\ \cdot \text{MAD}} \end{aligned} \tag{6.5}$$

6.2.6. PCA for composite weights

For the final performance composite, appropriate weights are necessary to combine all 3 pillar (z)scores. To that extent, a principal component analysis is performed that finds the linear combination of weights that maximises the shared variance of the 3 indicators. This way, pillar-level results are retained, but the composite reflects the dominant shared dimension across these correlated indicators of the same latent construct "performance". Table 6.3 shows the first component's loadings; with 72% of variance explained, this shows unidimensionality. The 3 pillars behave as one latent factor, and the normalised weights from the first component are appropriate for the final composite.

Loading (standardised)	PC1 (L2 norm)	PC1 (L1 norm)	Variance explained
Speed ($-\text{LRT}_{i\ell}$)	-0.591	0.342	
Efficiency ($-\text{F}_p$)	-0.604	0.349	72%
Sequence (SEQ_p)	-0.535	0.309	

Table 6.3: L2- and L1-normalised PC1 loadings from the PCA over the three pillars.

6.2.7. Performance composite

Finally, the standardised pillar scores, and the normalised PC1 loadings as weights w_s, w_p, w_q are used to form a performance composite $Perf_{i\ell}$ for participant i in level ℓ (see Equation 6.6d). Scoring can now be interpreted as: 0 typical performance for that level, +1/-1 about one robust SD above/below typical performance, respectively.

$$\text{Speed\_z}_{i\ell} = \text{z}_{\text{i}}^{\text{rob}}(-LRT_{i\ell}) \tag{6.6a}$$

$$\text{Efficiency\_z}_{i\ell} = \text{z}_{\text{i}}^{\text{rob}}(-F_{i\ell}) \tag{6.6b}$$

$$\text{Sequence\_z}_{i\ell} = \text{z}_{\text{i}}^{\text{rob}}(SEQ_{i\ell}) \tag{6.6c}$$

$$\text{Perf}_{i\ell} = w_s\text{Speed\_z}_{i\ell} + w_p\text{Efficiency\_z}_{i\ell} + w_q\text{Sequence\_z}_{i\ell} \tag{6.6d}$$

6.3. Performance Across Groups and Phases

The participant performance across groups and phases is analysed. For all subsequent sections, references to ”Intervention phase” and ”Main phase” correspond to the Intervention with the two different navigation control groups (Drag, Buttons) and the Holonomy VR environment, respectively. References to ”*PhaseX*”, where *Phase*:{Intervention, Main} and *X*:{1,2} correspond to levels 1 or 2 of the respective phase. In all the following visualisations, the group encoding by colour was kept consistent, with Buttons: ■, and Drag: ■.

6.3.1. Overall performance

To obtain a single score for the Intervention and Main phases, per participant, their performance scores are aggregated into one score per phase. Table 6.4 shows descriptive statistics for the mean performance for each group and each phase. A few notable mentions from it: the worst and best (averaged) runs in both phases were from the Buttons group. The mean performance in the Main phase by the Drag group was quite a bit higher than the Buttons group; the same holds for overall (Intervention and Main) performance. Figure 6.3 helps visualise this in more detail with boxplots for each of the levels separately. The two groups show similar performance in the Intervention, with the Buttons group outperforming Drag by a small margin. In the Main phase and overall, the Drag group outperforms Buttons by a larger margin.

Phase	Group	N	Mean	SD	Min	P25	P50	P75	Max
Intervention	Buttons	14	−0.169	0.899	−1.958	−0.517	−0.118	0.567	1.052
	Drag	14	−0.209	0.819	−2.000	−0.392	−0.032	0.408	0.723
Main	Buttons	14	−0.460	1.097	−2.138	−1.468	−0.254	0.043	1.667
	Drag	14	0.154	0.668	−1.003	−0.469	0.257	0.669	1.281
Both phases	Buttons	14	−0.315	0.784	−1.561	−0.794	−0.272	0.061	1.231
	Drag	14	−0.027	0.530	−0.874	−0.347	−0.205	0.240	1.002

Table 6.4: Descriptive statistics for mean performance (Perf) by phase and group. Quartiles are P25, P50 (median), P75.

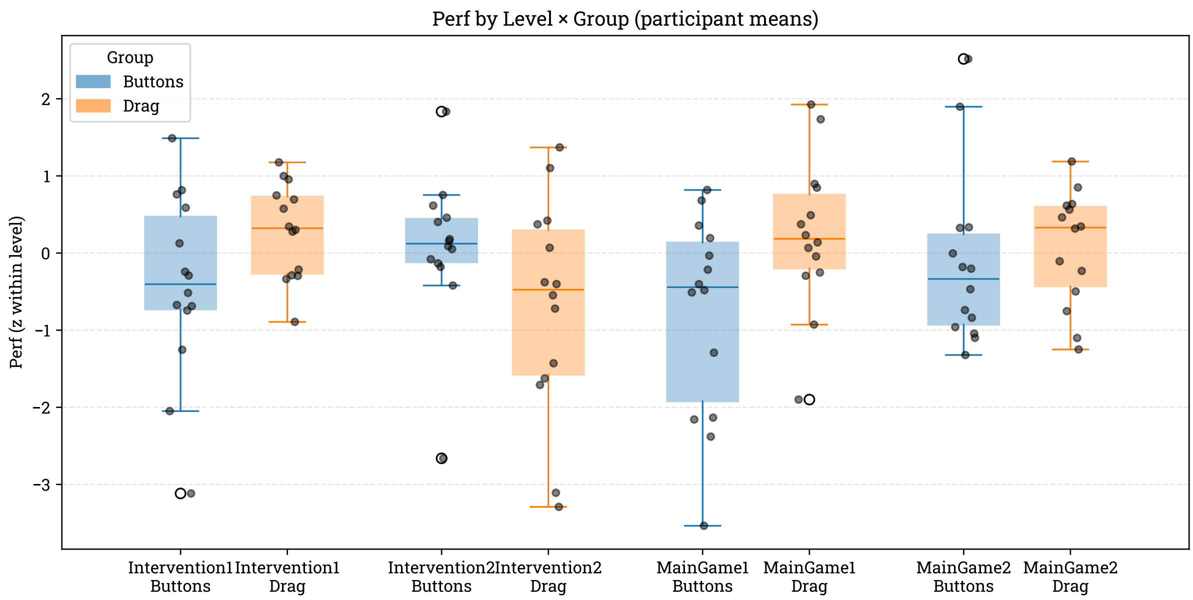


Figure 6.3: Mean performance per participant for all levels across groups.

6.3.2. Intervention vs. Main performance

Here, performance is compared across groups for each of the phases separately. Figure 6.4 shows violin plots of the full distribution of participants' mean performance by phase with width encoding density. In Intervention, the two groups' centres are similar, whereas in Main, the Drag violin is shifted upward with less overlap with Buttons, previewing the group difference.

Figures 6.5 and 6.6 show in more detail boxplot visualisations of the mean performance per group for each of the Intervention and Main phases, respectively. It can be seen in Intervention, medians and IQRs largely overlap across groups; in Main, the Drag median and IQR are shifted relative to Buttons, indicating better typical performance and a distributional shift, not just a few high performers.

Figures 6.7 and 6.8 show kernel density estimations for the performance distributions on a common scale, overlaid for both groups in the Intervention and Main phases, respectively. Intervention curves substantially overlap, suggesting comparable performance; in Main, the Drag density is shifted to higher values with a thinner left tail, consistent with a group advantage.

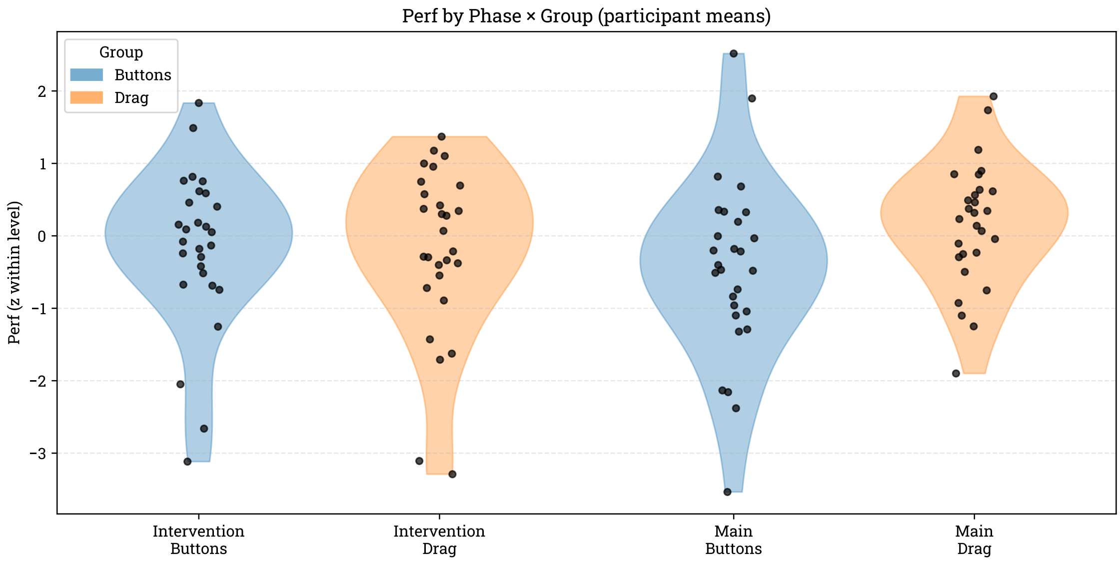


Figure 6.4: Violin plots for mean performance per participant across phases for both groups.

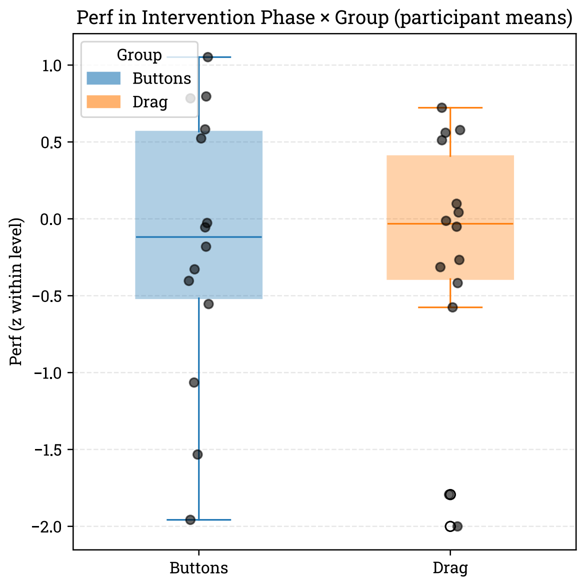


Figure 6.5: Mean performance per participant in the Intervention phase across groups.

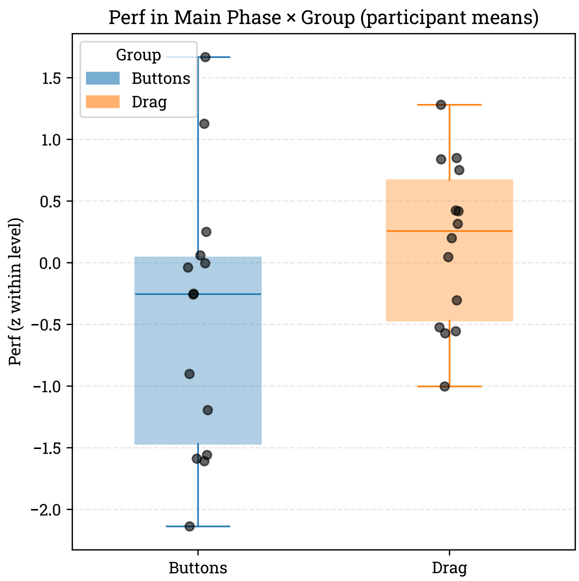


Figure 6.6: Mean performance per participant in the Main phase across groups.

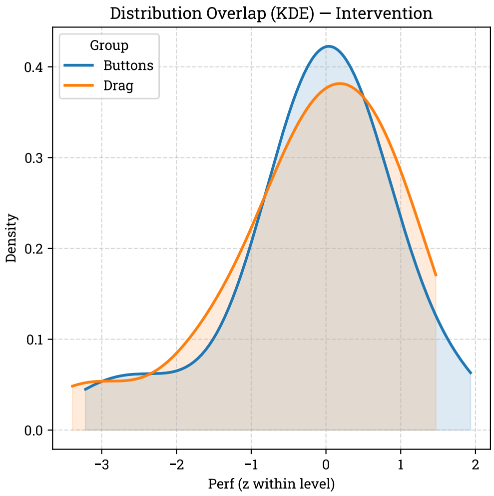


Figure 6.7: Kernel density estimates for Intervention performance distributions for both groups.

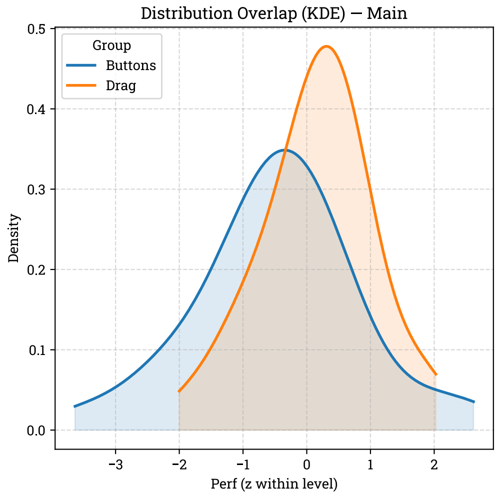


Figure 6.8: Kernel density estimates for Main performance distributions for both groups.

6.3.3. Intervention vs. Main pillar scores

It is also interesting to compare the pillar scores for each group across phases. Figures 6.9 and 6.10 summarise each group's pillar profile on a common, zero-centred scale (mid-ring = 0, outward = better). They can be read by comparing the distance to the outer ring on each spoke. Greater radial separation between groups on a spoke indicates a larger pillar gap. In Intervention, the polygons largely overlap, suggesting similar profiles across groups. In Main the Drag polygon is consistently farther out on all spokes, indicating a broader advantage rather than a single pillar dominance.

Figure 6.11 shows grouped bar plots of the pillar contributions to the mean performance per group across all phases. This view pinpoints which pillars drive the between-group difference and whether gaps are uniform or concentrated. With an axis-aligned comparison, it is clearer where the Drag advantage lies, especially for the Main phase. The Drag group had considerably better Speed and Sequence scores and a better Efficiency score. In the Intervention, the pillar gaps between groups were much closer. This aligns with the glyphs but gives a more quantitative sense of magnitude. Figure 6.12 shows a stacked bar plot. The bars decompose each group's mean performance into pillar contributions (to the total) encoded into segment lengths. The total bar length matches the overall composite. The figure visualises nicely each group's mean performance with its pillar contributions and complements the grouped bars. The latter isolates per-pillar gaps, while this figure shows how those gaps add up to the observed composite difference. Only the Drag group in the Main phase achieved a positive mean performance, with the Button's group performance being substantially inferior (and overall negative). The mean performance for both groups in the Intervention was comparable, but it is easy to pinpoint where the (small) difference lies, namely in the Speed and Efficiency pillars.

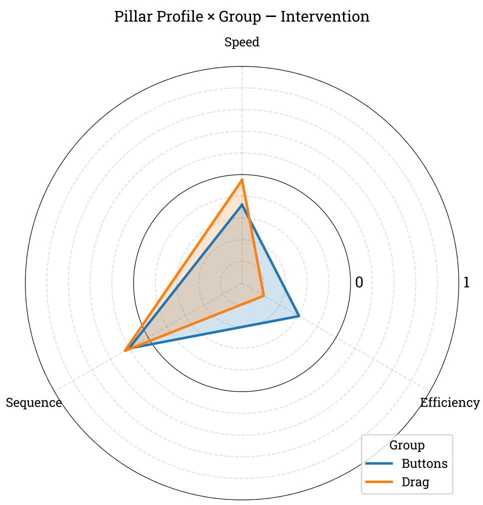


Figure 6.9: Mean pillar scores in the Intervention phase overlaid for both groups.

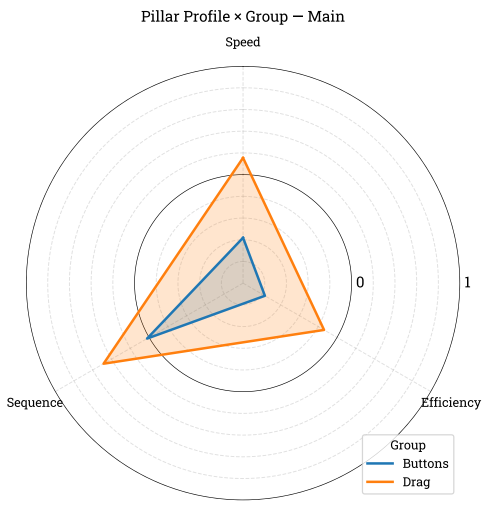


Figure 6.10: Mean pillar scores in the Main phase overlaid for both groups.

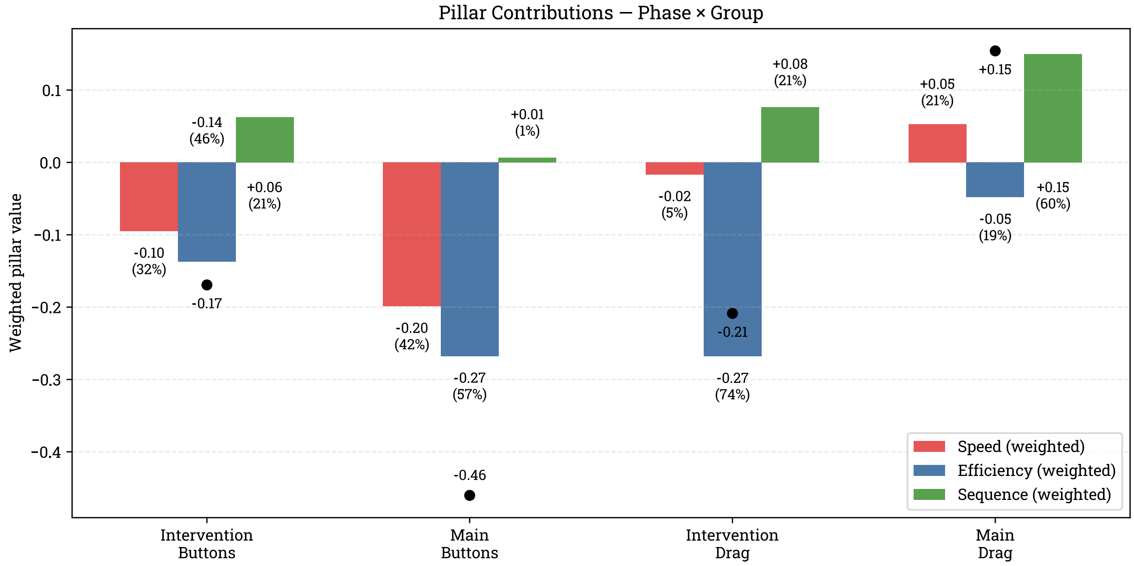


Figure 6.11: Grouped pillar contributions to mean performance per phase for both groups.

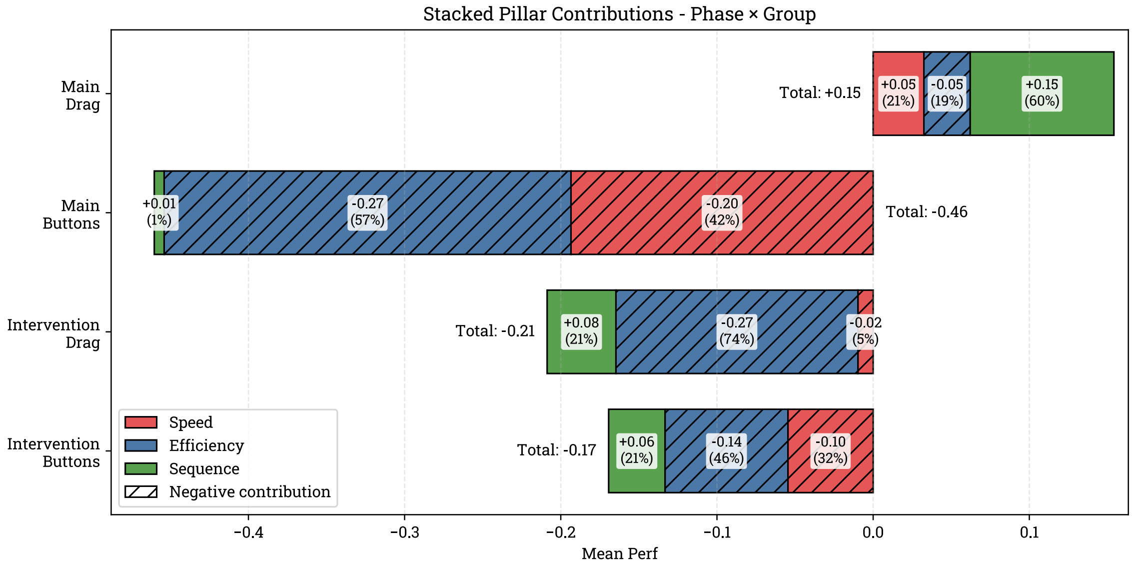


Figure 6.12: Stacked pillar contributions to mean performance per phase for both groups. Hatched shows a negative contribution, and bar length shows mean performance.

6.4. Main Phase Performance Analysis

The participants' performance in the Main phase is of particular interest since that is the part of the study where everyone was on equal footing. The levels, equipment, and environment were identical across groups; the distinguishing factor is the method of interaction they had in the Intervention phase.

6.4.1. Assumption checks for mean comparison

The Main performance data are checked for normality and homogeneity of variances. A simple ordinary least squares (OLS) model was fitted on the Main mean performance per participant to check for normality. Figures 6.13 and 6.14 show the Q-Q and histogram plots for the residuals of the model. The point spread on the Q-Q plot is pretty narrow, and the residuals seem to be normally distributed. The same goes for the histogram being roughly symmetric and centred at 0.

A Shapiro-Wilk test and a Levene's test were also run on the Main mean performance, and the results can be seen in Table 6.5. The Shapiro-Wilk test statistic W measures how close the sample distribution is to normal, with 1 indicating closer agreement. The p-value tests the null-hypothesis of normality. Both p-values are well above $\alpha = 0.05$ and W is close to 1, so normality can be assumed for either group. This is consistent with the Q-Q plot and residual histogram. The Levene's test checks for homogeneity of variances across groups. Its statistic W (or F) evaluates the null that the group variances are equal. With p-value > 0.05 again, equal variances can be assumed. The checks support using a parametric mean comparison.

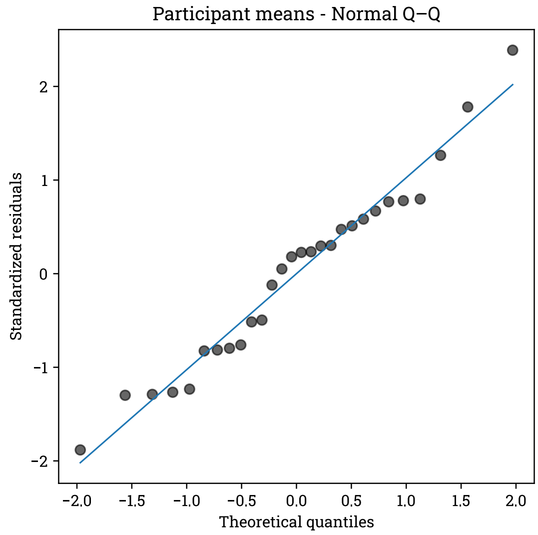


Figure 6.13: Q-Q plot of residuals from a simple model fit on Main mean performance per participant.

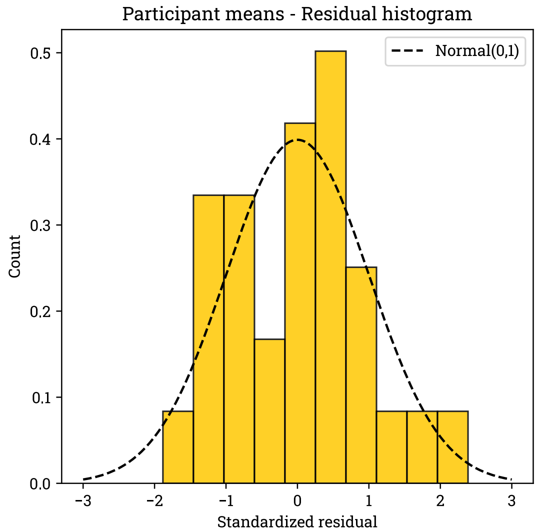


Figure 6.14: Histogram of residuals from a simple model fit on Main mean performance per participant.

Test	Group	W	p-value
Shapiro-Wilk	Buttons	0.953	0.606
	Drag	0.963	0.776
Levene's	-	2.415	0.132

Table 6.5: Results of the Shapiro-Wilk and Levene's test on the mean performance in the Main phase.

6.4.2. Between-group tests

Since normality and homogeneity of variances are established, a (Student's) t-test can be used to compare mean performance for the Main phase across groups at the participant level. A Mann-Whitney U test can also be used as a non-parametric alternative. The results of both of these tests, along with Cohen's d to provide an effect size and CIs, can be found in Table 6.6.

The two-sample t-test asks whether the group means differ. Its statistic ($t = 1.790$) is the observed mean difference scaled by its standard error. The two-sided $p = 0.085$ is the probability of seeing a value at least this extreme if the true mean difference were zero. At $p > 0.05$, the equal means at 5% is not rejected, though the value is suggestive rather than negligible. The Mann-Whitney U test checks whether the two distributions are identical based on the rank ordering of all observations. Here $U = 135.000$ with $p = 0.094$, essentially mirroring the t-test. There is a similar trend, but not conventionally significant at $\alpha = 0.05$.

To gauge magnitude, Cohen's d standardises the mean difference in units of the pooled standard deviation. The point estimate $d = 0.677$ corresponds to a moderate difference (about two-thirds of a SD) in favour of the group with the higher mean (Drag, per the descriptives). However, the 95% confidence interval $[-0.037, 1.666]$ includes zero, meaning at $N = 28$, the uncertainty remains wide.

Test	t-value	U-statistic	p-value	d	95% CI
(Student's) t-test	1.790	–	0.085	–	–
Mann–Whitney U	–	135.000	0.094	–	–
Cohen's d	–	–	–	0.677	[−0.037, 1.666]

Table 6.6: Group comparison tests for mean performance in the Main phase.

To complement the tests mentioned above, a mixed-effects model is used. The model is similar to the OLS model that was fit for the residuals, but also accounts for between-level difficulty:

$$\mathrm{Perf}_{i\ell} = \beta_0 + \beta_1 Drag_i + u_{participant_i} + u_{level_\ell} + \varepsilon_{i\ell} \tag{6.7}$$

The two-sample tests above compare group means on a per-participant summary and ignore two things: repeated runs per participant and systematic differences in level of difficulty. The mixed-effects model retains all runs, models Drag vs. Buttons as a fixed effect, and includes random intercepts for participant and level to prevent within-person correlation from inflating the evidence. Additionally, harder/easier levels are absorbed rather than attributed to the group effect. Here β_1 is the mean difference $Drag - Buttons$ on the performance scale after adjusting for participant and level.

The naive Wald test for β_1 was anti-conservative, estimating participant random variance near zero and effectively treating the 56 runs as independent. To respect the data structure, cluster-aware permutations ($N = 2000$) are used to shuffle group labels per participant. The observed fixed effect was $\hat{\beta}_1 \approx 0.614$ with a permutation $p = 0.092$. This means that on average, Drag scores about 0.61 units higher than Buttons after adjusting for level and clustering. The cluster permutation $p = 0.092$ indicates suggestive but not conventionally significant evidence at $\alpha = 0.05$. To obtain a confidence interval for $\hat{\beta}_1$, participants are similarly resampled with replacement ($N = 2000$) to get a bootstrap distribution. This yielded a 95% CI $[0.107, 1.126]$ which broadly agrees with the direction and magnitude seen in the t/Mann-Whitney results. The CI excludes 0, implying significance at 5% but permutation $p \approx 0.09$. Small discrepancies are common at this sample size, so this is interpreted conservatively.

Across methods that assume normality and equal variances (two-sample t), rank-based distribution-free (Mann-Whitney), and model the full repeated-measures structure (mixed-effects model), the story is consistent. Drag tends to outperform Buttons by a moderate amount, but with the present sample, the evidence is not definitive at 5%. Reporting the point estimate and its uncertainty (permutation p, bootstrap CI) is the most informative summary. A larger N would be needed to narrow the CI and turn this suggestive result into a more precise one.

6.4.3. Controlling for baseline (Intervention) performance

Because the Intervention phase used different control schemes by design (Buttons vs Drag), the primary inference thus far focused on the Main phase. As a complementary analysis, an ANCOVA is fit at the participant level. It essentially asks whether groups differ in Main performance at a common baseline level. $\mathrm{Drag}_i = 1$ is coded for Drag and 0 for Buttons:

$$MainPerf_i \;=\; \beta_0 \;+\; \beta_1\,\mathrm{Drag}_i \;+\; \beta_2\,InterventionPerf_i^{\text{centred}} \;+\; \varepsilon_i, \tag{6.8}$$

The adjusted Drag effect remained positive: $\hat{\beta}_1 = 0.620$ (SE $= 0.350$), 95% CI $[-0.103,\ 1.343]$, $p = 0.090$, indicating that holding baseline constant, Drag tends to score higher in Main. Once again, CI crosses zero, so this is suggestive but not definitive, but it is consistent in direction with the mixed-model results. The baseline (Intervention) slope was small and not significant: $\hat{\beta}_2 = 0.279$ (SE $= 0.286$), 95% CI $[-0.312,\ 0.869]$, $p = 0.339$, implying that Intervention performance carries limited predictive information for Main once group is accounted for. This is no surprise, given the different control schemes used in Intervention. Adding an interaction term $group \times baseline$ had a negligible incremental fit ($\Delta R^2 \approx 0.015$; AIC/BIC worsened; $F(1, 24) = 0.41, p = 0.527$). This supports the ANCOVA assumption that the baseline-Main relationship is parallel across groups. Overall model fit was modest, $R^2 = 0.129$, indicating that group and baseline together only explained $\approx 13\%$ of the variance in Main performance.

Adjusting for Intervention performance does not change the qualitative conclusion: Drag tends to outperform Buttons in the Main phase. Because Intervention used different control schemes, this ANCOVA is presented as an adjusted, supportive analysis rather than a causal "change-from-baseline" model.

6.4.4. Performance delta across phases

As an exploratory cross-check, it is worth looking at a per-participant change score Δ_i and comparing between groups:

$$\Delta_i = \overline{MainPerf_i} - \overline{InterventionPerf_i} \tag{6.9}$$

The change score Δ_i captures within-participant trajectories from Intervention to Main, asking whether Drag improved more than Buttons when each person serves as their own benchmark. Figure 6.15 presents Δ_i as a waterfall plot; bars above zero indicate improvement and overlaid group means summarise central tendency. This view makes clear the heterogeneity of responses (not all participants move in the same direction) while showing that Drag's average change is more positive than Buttons.

$\bar{\Delta_D}$ for the Drag group was 0.363, and $\bar{\Delta_B}$ for the Buttons group was -0.291, yielding a between-group difference of 0.654. A Welch's t-test resulted in $t = 1.497$, $p = 0.147$, and a Mann-Whitney U test resulted in $U = 116.0$, $p = 0.421$. A participant-label permutation test gave $p = 0.152$. Directionally, Drag improved from Intervention to Main while Buttons declined on average. This is consistent with the Main phase analysis, but at $N = 14$ per group, evidence is suggestive and not statistically conclusive.

Given the different control schemes used in the Intervention by design, Δ is a post-randomisation contrast without an equivalent baseline. The change score is also noisy since the baseline reliability is modest, with $\hat{\beta}_2 = 0.279$, $p = 0.339$ as seen in Subsection 6.4.3; hence, this is treated as descriptive support rather than a stand-alone causal estimate.

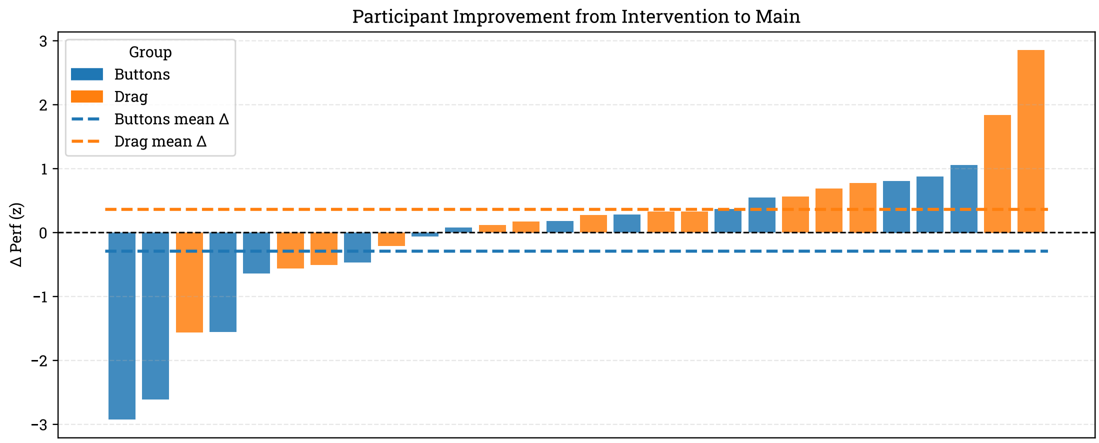


Figure 6.15: Performance delta from Intervention to Main per participant for both groups.

6.5. Exploratory Analysis

Here, other facets of the study are looked at to gain more insight into additional relationships that may exist among the data. These outcomes are to be treated as descriptive and/or exploratory; they are not intended for inference or causality investigation.

An understanding score $USC \in [0, 4]$ is formed, which reflects the correct participant answers in the "Hyperbolic World Understanding" part of the survey as explained in Subsection 5.4.3. A point is awarded for every correct answer to the 4 multiple-choice questions. Table 6.7 shows the mean $\bar{USC}$ score for each group and overall, along with the correlation values for USC and overall performance (both phases). USC shows no relationship with overall performance. A Welch's t-test with $p = 0.114$, and a Mann-Whitney U test with $p = 0.169$, further confirm it since p-values are above the significance level. A relationship between USC and the participants' field of study was also examined, but with sparse, uneven distributions, no conclusive evidence was found.

Group	N	$\bar{USC}$ score	SD	Spearman's ρ	p-value
Buttons	14	2.643	0.842	0.092	0.756
Drag	14	2.071	0.997	0.052	0.861
Overall	28	2.357	0.951	0.002	0.993

Table 6.7: Mean understanding score $\bar{USC}$ per group and Spearman's ρ correlation of USC with overall performance.

Additionally, the predictive validity across phases was examined using Spearman's ρ between the Intervention mean performance and the Main mean performance. This was further extended to (z-standardised) pillar-wise correlations as well. As seen in Table 6.8, correlations are positive but small and imprecise with large p-values. This is consistent with the observed modest reliability of single-session baselines.

Group	Perf		Speed		Efficiency		Sequence	
	ρ	p-value	ρ	p-value	ρ	p-value	ρ	p-value
Buttons	0.244	0.401	0.389	0.169	0.209	0.474	0.323	0.260
Drag	0.196	0.503	0.174	0.553	0.143	0.626	0.314	0.274

Table 6.8: Correlation between Intervention and Main phase scores (at the participant level) for performance (Perf), and all pillars individually (Speed, Efficiency, Sequence).

To further look into performance differences per group from the Intervention to the Main phase, participants were stratified within group into Low/Mid/High based on their Intervention performance, so each comparison is made within the same baseline band. The stratification asks whether the $Drag-Buttons$ difference from Intervention→Main is merely a by-product of baseline composition (e.g. Drag having more high baselines) or whether it persists across baseline levels. The mean performance per group was calculated and compared using a Mann-Whitney U test, reported alongside Hedges' g (smaller $N = 14$) to obtain an effect size. The results shown in Table 6.9 favour the Drag group in all strata, since the Drag-Button difference is positive, largest for the Low stratum ($g = 1.302$, $p = 0.095$), and smaller, imprecise in Mid/High ($g = 0.379; 0.429$, $p = 0.680; 0.690$). This pattern indicates that Drag's advantage in Main is not explained by Drag participants starting higher in Intervention. On the contrary, the advantage was numerically largest among those who performed lower at baseline. This, combined with the ANCOVA in 6.4.3, where the baseline slope and $group \times baseline$ interaction were small, suggests limited baseline carry-over. Given the very small per-stratum N, the focus is on effect-size direction and consistency over p-values. This result is treated as exploratory support for the primary Main-phase finding.

Strata	Buttons	Drag	Drag - Buttons	MWU U	MWU p	Hedges' g
Low	−0.796	0.053	0.849	21.000	0.095	1.302
Mid	−0.404	−0.099	0.305	10.000	0.686	0.379
High	−0.169	0.458	0.627	15.000	0.690	0.429

Table 6.9: Differences in performance from Intervention to Main, stratified by Intervention performance.

6.6. Interaction and User Experience Analysis

The interaction analysis concerns only the Main phase, since that is where the interactive environment elements were introduced to the participants.

For each run in the Main phase, a tree interaction I_{Tree} timer kept track of the total duration of the participant's interaction with trees in the Holonomy VR environment. Similarly, a hedge interaction I_{Hedge} timer kept track of the total duration of interaction with the hedge in Holonomy VR. A sum of the quantities I_{total} (see Equation 6.10a) captures the entire duration a participant spent interacting with either trees or the hedge across their runs in the Main phase. Since the distribution is right-skewed with zeroes (not all participants interacted with the environment), a $log(1p)$ transform and z-score standardisation are used to obtain zI_{total} (see Equation 6.10b). The total time spent in the Main phase zT_{Main} is also calculated (see Equation 6.10d) by summing the time across runs and then standardising in the same way to obtain zI_{total}.

$$\mathrm{I}_{\mathrm{total}} = I_{Tree_1} + I_{Hedge_1} + I_{Tree_2} + I_{Hedge_2} \tag{6.10a}$$
$$\mathrm{zI}_{\mathrm{total}} = z(\log(1 + I_{total})) \tag{6.10b}$$
$$\mathrm{T}_{\mathrm{Main}} = T_{Main_1} + T_{Main_2} \tag{6.10c}$$
$$\mathrm{zT}_{\mathrm{Main}} = z(\log(1 + T_{Main})) \tag{6.10d}$$

The 26 User Experience Questionnaire (UEQ) items are transformed on a $[-3, 3]$ scale, with negative values attributed to the negative adjective of each item. The items are mapped to the 6 official scales by the UEQ's authors: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty. Mean scores for each scale are obtained per participant. The overall scale means for the responses regarding the Main phase can be seen in Table 6.10 for completeness. At this sample size, confidence intervals are relatively wide, so estimates should be read as directional rather than precise.

To contextualise magnitudes, Figure 6.16 overlays the per-scale means on the UEQ benchmark bands (as per the authors). The coloured bands indicate percentile ranges from the UEQ reference, and the black markers the obtained Main-phase means. On this scale, the Main phase was evaluated as Above-Average/Good/Excellent on all dimensions except for Perspicuity, which achieved a Below-Average score. This is understandable given the learning demands of non-Euclidean navigation and first-time exposure. Stimulation and Novelty are comparatively high, suggesting the experience is engaging and distinctive. Efficiency and Dependability cluster around Above-Average/Good, indicating that perceived effectiveness and reliability were generally positive despite the elements of unfamiliarity.

The benchmark is treated as a comparison in a descriptive context rather than a hypothesis test. This situates the achieved scale means against a broad UEQ reference but does not adjust for domain differences or the sample size. The corresponding table and graph for the Intervention phase scales can be found in the Appendix A.

Scale	Mean	SD	Confidence	95%CI
Attractiveness	1.702	1.117	0.414	[1, 289, 2.116]
Perspicuity	0.973	1.220	0.452	[0, 521, 1.425]
Efficiency	1.250	1.145	0.424	[0, 826, 1.674]
Dependability	1.232	1.078	0.399	[0, 833, 1.631]
Stimulation	1.938	0.778	0.288	[1, 649, 2.226]
Novelty	2.063	0.648	0.240	[1, 823, 2.302]

Table 6.10: UEQ scale means and descriptives from all participants for the Main phase.

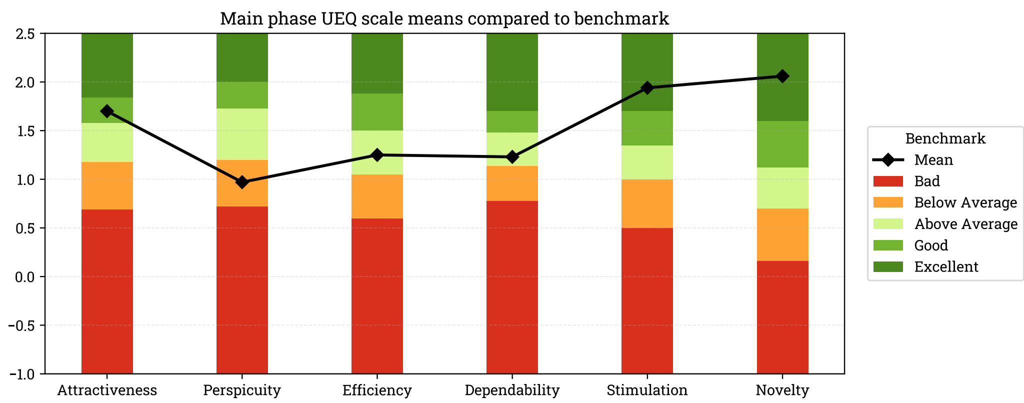


Figure 6.16: The UEQ scale means from the Main phase responses compared to a benchmark.

As discussed in Subsection 5.4.3, the survey includes two self-reported multiple-choice questions, which will be referred to as MCQ5 and MCQ6, that regard interactions in the Holonomy VR environment. MCQ5 reflects the perceived time spent interacting, while MCQ6 reflects how memorable or engaging the interactions were. Both are on a 6-point scale, with category 6 = ”chose not to / not applicable” respectively. For the analysis, 6 was excluded, and the remaining 1-5 were standardised to $zMCQ5$ and $zMCQ6$ (where necessary).

6.6.1. Objective interaction vs. self-reports

An ordinal logistic model was fit to check if the objectively measured time zI_{total} captures what participants experienced through the self-reports by MCQ5 and MCQ6. Table 6.11 shows the odds ratio (per 1 standard deviation in log interaction time) and Spearman's ρ between the self-reported responses and the log interaction time. At $OR = 1.931$ for MCQ5, the odds of one choosing a higher MCQ5 category are ~93% greater for someone one SD higher in objective time (95% CI $[0.926, 4.027]$, $p = 0.079$). This is paired with Spearman's ρ as a model-free, rank-based robustness check, which for MCQ5 gave $\rho = 0.404$, at $p = 0.041$. MCQ6 is directionally positive ($OR = 1.684$; $\rho = 0.138$), but imprecise at $N = 28$. Together, these results suggest the log-time measure has reasonable correspondence with perceived time, while engagement/memorability is not tightly determined by time alone. This is consistent with later findings (see Subsection 6.6.2) where MCQ6 relates more strongly to UEQ scales than to logs. This is treated as a validity check, not a causal claim.

Survey Question	Odds Ratio	95%CI	p-value	Spearman's ρ	Spearman's p-value
MCQ5	1.931	$[0.926, 4.027]$	0.079	0.404	0.041
MCQ6	1.684	$[0.846, 3.355]$	0.138	0.187	0.341

Table 6.11: Odds ratio per +1 SD in log interaction time and Spearman's ρ for MCQ5 and MCQ6 and zI_{total}.

6.6.2. Interaction vs. UEQ scales

Interaction can be compared to the 6 UEQ scales through 3 lenses: objective interaction time, MCQ5 (perceived interaction time), and MCQ6 (perceived engagement). An ordinary least squares model with robust SEs was fit for each UEQ scale [-3,3] on each of the standardised zI_{total}, $zMCQ5$, and $zMCQ6$. Since 6 scales are tested, the false discovery rate (FDR) is controlled across p-values at $q = 0.10$. Spearman's ρ is also calculated for each scale as a rank-based robustness check. The OLS coefficient β gives the expected change in a UEQ scale $[-3, 3]$ per +1 SD in the tested metric.

Seeing the results in Table 6.12, MCQ6 was found as best, showing its strongest, FDR-significant association with Efficiency ($\beta = 0.766$, $\rho = 0.627$, $p < 0.001$). This means a participant one SD higher on MCQ6 is, on average, 0.8 scale points higher on Efficiency. Dependability ($\beta = 0.489$, $p = 0.091$, $\rho = 0.326$, $q = 0.092$) and Stimulation ($\beta = 0.369$, $p = 0.047$, $\rho = 0.359$, $q = 0.093$) are positive and near-threshold after FDR. Attractiveness and Perspicuity trend positively but are imprecise at $N = 28$; Novelty is small. In models adjusting for total interaction time zI_{total} and total time spent in Main zT_{Main}, Efficiency remains robust while other coefficients slightly diminish, indicating the MCQ6-UEQ link is not explained by exposure time alone. The full results obtained from all models can be found in Appendix A.

Objective time had small UEQ effects, whereas perceived engagement aligns with perceived quality (especially Efficiency). This is a pattern that matches the item-level findings (see 6.6.3), and supports using MCQ6 as the primary perceptual driver in the second research question.

UEQ Scale	β	95%CI	p-value	FDR q	Spearman's ρ	Spearman's p-value
Attractiveness	0.538	$[-0.067, 1.144]$	0.081	0.122	0.409	0.031
Perspicuity	0.436	$[-0.097, 0.968]$	0.109	0.130	0.275	0.157
Efficiency	0.766	$[0.392, 1.140]$	0.000	0.000	0.627	0.000
Dependability	0.489	$[0.046, 0.932]$	0.031	0.092	0.326	0.091
Stimulation	0.369	$[0.006, 0.733]$	0.047	0.093	0.359	0.061
Novelty	0.195	$[-0.061, 0.452]$	0.136	0.136	0.268	0.169

Table 6.12: Results of OLS regression model for each UEQ scale on $zMCQ6$ with Spearman's ρ and FDR control.

6.6.3. Interaction vs. UEQ items

Similarly, interaction can be compared to the individual UEQ items to ask which specific adjectives drive the scale-level effects. Meaning, is the MCQ6-UEQ link broad, or concentrated in particular facets of the questionnaire? For each UEQ item $[-3, 3]$, the item score is regressed on $zMCQ6$, with β being the expected item shift (in scale points) per +1 SD in MCQ6. Multiplicity is again controlled with FDR at the same threshold $q = 0.10$ and Spearman's ρ as a rank-based check.

The (near) FDR-significant positives, as seen in Table 6.13, cluster in Efficiency (items: 20 [inefficient/efficient], 23 [cluttered/organised], 22 [impractical/practical]), with additional support in Dependability (11 [obstructive/supportive]), Stimulation (5 [demotivating/motivating]), Novelty (26 [conservative/innovative]), and Attractiveness (16 [unattractive/attractive]). Effects are sizeable ($\approx 0.6; 1.0$ points per +1 SD MCQ6), and Spearman's ρ agrees in sign and rank. This indicates that higher perceived engagement/memorability aligns with feeling more efficient, organised, and supported, with secondary feelings of more motivation, innovation, and attractiveness. This mirrors the scale-level finding where Efficiency is the clearest responder to MCQ6.

At the scale-level, objective time zI_{total}, and perceived time $zMCQ5$ showed small, non-significant associations after FDR. In exploratory item fits, they similarly produced no coherent FDR-robust pattern at $N = 28$. To avoid accumulation of marginal tests and false positives, the MCQ6→items results are placed in the foreground. The full table with all items can be found in the Appendix A for transparency. This keeps the item analysis aligned with the main second research question signal of engagement↔perceived quality.

UEQ Scale	Item #	β	95% CI	p-value	FDR q	Spearman ρ	Spearman p
Attractiveness	16	0.858	[0.186, 1.529]	0.012	0.054	0.522	0.004
Perspicuity	13	0.588	[−0.036, 1.212]	0.065	0.187	0.344	0.073
Efficiency	20	0.912	[0.468, 1.356]	0.000	0.001	0.550	0.002
	23	1.013	[0.448, 1.578]	0.000	0.006	0.495	0.007
	22	0.757	[0.170, 1.344]	0.012	0.054	0.450	0.016
Dependability	11	0.654	[0.173, 1.135]	0.008	0.050	0.391	0.040
Stimulation	5	0.634	[0.119, 1.150]	0.016	0.059	0.448	0.017
Novelty	26	0.509	[0.173, 1.135]	0.002	0.014	0.478	0.010

Table 6.13: (Near) FDR-significant UEQ items of the OLS regression model fit on $zMCQ6$ with Spearman's ρ.

7
Discussion

This chapter interprets the empirical findings with respect to the two research questions and situates them in the broader literature. The first research question is revisited, combining distributional evidence, mixed-effects inference, and sensitivity checks to assess whether Drag confers a Main-phase advantage under matched conditions. Then the second research question is addressed, contrasting objective interaction time with self-reports and showing how perceived engagement (MCQ6) relates to UEQ scales and items, including time-adjusted models and FDR control. Afterwards, key methodological considerations are laid out that guided inference, like choice of estimand, dependence handling, and robustness checks. This is followed by limitations and alternative readings regarding statistical power, zero-inflation, measurement scope, and shared-method variance. The implications for design and theory are then presented, and subsequently, the findings are positioned within the state of the art. This is clarified by how they extend prior work and where future studies should take the platform next.

7.1. Performance in Context

The first hypothesis essentially posits that participants trained with the Drag interaction method ultimately perform better than participants trained with the Buttons interaction when both are faced with the same conditions in the Main phase. The outcome is a composite performance score $Perf$, derived from three pillars: Speed, Efficiency, and Sequence. Each pillar was standardised within level and combined in a way that mirrors the dominant component from a PCA that explained 72% of variance over the pillars. This yields a scale where 0 reflects level-typical performance and ± 1 approximately one robust standard deviation from typical performance. That construction matters for interpretation: a between-group difference in $Perf$ can be read as a standardised advantage in overall task fluency that combines timing, spatial optimality, and action ordering.

When the two Main runs were averaged per participant (the primary estimand for the first research question), the Drag group outperformed the Buttons group by roughly 0.61 $Perf$ points, roughly two-thirds of a robust SD. Classical tests agreed on direction and magnitude and produced p-values just above conditional thresholds with (Student's) $t = 1.790$, $p = 0.085$, and Mann-Whitney $U = 135$, $p = 0.094$ with the effect size by Cohen's $d = 0.677$ 95% CI$[-0.037, 1.666]$. Assumption checks supported parametric testing: Shapiro–Wilk within groups $p \geq 0.606$; Levene $p = 0.132$, but at $N = 14$ per group, these should not be over-interpreted. With few participants per group, the p-values from the tests reflect limited power rather than conflicting signals. All point estimates place Drag higher than Buttons by an amount that would be practically noticeable in the task, albeit just above conventional thresholds and with a wide confidence interval.

Because runs are nested within participants and levels differ in difficulty, a run-level model is also fit with random intercepts for participant and level. In maximum-likelihood estimation, the participant variance collapsed toward zero, which made the usual Wald p-values anti-conservative. To restore valid inference without abandoning the model's structure, a cluster-aware label permutation at the participant level is used. Group labels were shuffled by person, refitting the model, and situating the observed β in the empirical null distribution. The observed group effect from the mixed model was again $\hat{\beta} = 0.614$, 95% CI $[0.107, 1.126]$ with a permutation $p = 0.092$. This agreement with the participant-level analysis indicates that the Main phase Drag advantage is not an artefact of treating runs as independent or of single-level characteristics.

A natural follow-up is whether the performance difference in the Main phase persists after accounting for initial ”ability” as expressed in the Intervention phase. An ANCOVA that regressed Main performance on group and Intervention performance gave an adjusted Drag coefficient $\hat{\beta}_1 = 0.620$, 95% CI $[-0.103, 1.343]$, $p = 0.090$ with a modest overall $R^2 = 0.129$. The slope on Intervention was small and imprecise $\hat{\beta}_2 = 0.279$, 95% CI $[-0.312, 0.869]$, $p = 0.339$. A $group \times baseline$ interaction was tested, which essentially asks ”do better baselines translate to better Main performance differently by group?”. The interaction did not improve model fit $\Delta R^2 \approx 0.015$, AIC/BIC worsened, $p = 0.527$. Two things follow from this. First, adjusting for baseline Intervention performance does not change the qualitative conclusion; if anything, it stabilises it by showing the group difference is not driven by an imbalance in baseline ability. Second, the lack of an interaction means the homogeneity of slopes assumption is reasonable in this dataset: the relationship between prior performance and Main performance looks similar across groups.

A simple per-participant change score is also computed $\Delta = Main - Intervention$, and Δ is compared between groups. The Drag group's change was larger, with a difference of 0.654, but the evidence was modest with a Welch's $t = 1.497$, $p = 0.147$ and Mann-Whitney $U = 116.0$, $p = 0.421$. This is consistent in direction with the Main-only contrast, but is softer statistically. This is unsurprising given the design of the experiment; the Intervention deliberately differed across groups, and this is why the Main-only estimand is the most faithful to the research question.

In sum, the entirety of the first hypothesis analyses point in the same way: the Drag group carries a Main-phase performance advantage of moderate size, stable across analytical lenses: participant-mean tests, dependence-aware mixed model with permutations, ANCOVA adjusting for Intervention. The effect's practical meaning is straightforward on the $Perf$ scale: Drag participants executed the Main task faster, along shorter and more appropriate paths and action orders by roughly two-thirds of a level-standard deviation. The statistical uncertainty is a function of sample size rather than mixed or contradictory evidence.

7.2. Engagement and User Experience

The second hypothesis examines how variation in user-driven interaction with reactive elements, trees and hedges that respond haptically, relates to user experience in the Main phase. Here, objective interaction time and self-reports are treated as complementary views of ”interaction intensity”. Objective time aggregates the seconds spent directly engaging with reactive elements across both runs in the Main phase. These counts are zero-inflated and right-skewed because not all participants interacted with the environment; therefore, the counts are transformed using $log1p$ and standardised with z-scores. Self-reports captured two related but distinct quantities: MCQ5, a self-estimate of how long participants interacted for; and MCQ6, a rating of how engaging or memorable those interactions felt. Both items offered a ”6 = chose not to / not applicable” category, which was excluded in level analyses and flagged as a special case in ordinal checks.

Before connecting interaction intensity to user experience, the two representations of intensity (objective time, MCQ5/MCQ6) were checked to see if they align with each other. Participants who had higher log-scaled objective time tended to report higher MCQ5 categories with ordinal-logit $OR = 1.931$ per +1 SD, 95% CI $[0.926, 4.027]$, $p = 0.079$, and Spearman $\rho = 0.404$. The association with MCQ6 was positive but imprecise with $OR = 1.684$, 95% CI $[0.846, 3.355]$, $p = 0.138$, and Spearman $\rho = 0.187$. The pattern suggests that objective time captures something participants notice and can self-estimate, and is at least directionally grounded. Perceived engagement, however, is not reducible to time alone and seems to tap a distinct dimension; this is an important conceptual point for the following interpretations. Again, with wide confidence intervals and moderate p-values, precision is modest at $N = 28$.

The primary concern in the second research question and hypothesis is the link from perceived interaction length/engagement (MCQ5/MCQ6, respectively) to user experience, which is measured by the UEQ scales and items. Across the 6 UEQ scales, the clearest and most robust association was between MCQ6 and Efficiency. In the unadjusted model, a one-SD increase in MCQ6 corresponded to about $\beta = 0.766$ points on the $[-3, 3]$ Efficiency scale (95% CI $[0.392, 1.140]$, $p < 0.001$). Crucially, that relationship remained strong after adjustment for both log objective interaction time and total exposure time in Main. The coefficient increased to $\beta = 0.811$ (95% CI $[0.288, 1.335]$, $p = 0.002$). This "specificity" check matters for interpretation: if the MCQ6-Efficiency link were merely a factor for "they stayed longer", it should diminish notably once time is in the model, but did not. The false-discovery-rate (FDR) correction, applied across the 6 UEQ scales to guard against family-wise cherry-picking, kept Efficiency as significant both before and after adjusting for time. The practical meaning is that participants who experienced the interactions as engaging or memorable also experienced the system as more efficient to use, and this relationship is not simply explained by raw exposure to the environment of Holonomy VR.

Two other scales, Dependability and Stimulation, also moved positively with MCQ6. Their unadjusted slopes were moderate, with coefficients $\beta_D = 0.489$, $\beta_S = 0.369$ (per +1 SD), and $p_D = 0.031$, $p_S = 0.047$ respectively. Adjusting for time, still kept the estimates positive with $p \approx 0.07$ and FDR-adjusted values near the reporting threshold ($q = 0.10$). Although these are statistically softer than Efficiency, they are theoretically coherent: feeling the interactions "click" can reasonably bring a sense that the system is reliable and that the experience is stimulating or exciting. The remaining scales: Attractiveness, Perspicuity, and Novelty showed smaller positive trends with wider intervals. These do not contradict the story but should be treated as descriptive in this sample.

To avoid hiding effects by averaging items into scales, item-level regressions of each response are also examined on MCQ6 with FDR control across the 26 tests. The significant items cluster exactly where the scale results suggest. Multiple Efficiency items were positively associated with MCQ6, and there were additional significant items under Dependability, Stimulation, Novelty, and Attractiveness. This granularity helps interpretability. It indicates that the engagement-UX link is not a single outlier item or an averaging artefact; rather, it is made up of specific points in perceived efficiency (e.g. being fast, organised, effective) and extends to a subset of reliability and hedonic items.

The full range of models in Appendix A show that objective interaction time (zI_{total}) or self-reported interaction time (MCQ5) do not survive the FDR control ($q \geq 0.915$) and are generally statistically weak; hence, the focus is shifted to MCQ6. This strengthens the interpretation that memorable/engaging interaction predicts perceived efficiency rather than any simple exposure-driven mechanism. The pattern fits the hypothesis that "learning by doing" and enactive exploration improve practical proficiency. Taking everything together, a chain is suggested with progressively stronger links as one moves from raw time to perceived engagement to UX: objective time (zI_{total}) → self-estimated time (MCQ5) → perceived engagement (MCQ6) → user experience. The first arrow is present but modest; the second and third are the points where the signal becomes clear. Conceptually, this favours an enactive perspective. It is less the sheer duration of touching reactive elements and more whether those touches yielded meaningful sensorimotor coupling that shaped how the system felt to use.

The evidence supports the second hypothesis in its perceptual core; participants who experienced the reactive interactions as engaging reported a better user experience. A plausible alternative explanation could be common-source variance since both MCQ6 and UEQ are self-reports collected post-treatment. A positive effect might inflate correlations, which is why the time-adjusted analysis helps here. Moreover, the absence of MCQ5 to UEQ effects, despite MCQ5 being a similar kind of self-report, suggests the MCQ6 link is not merely a result of global positivity but pulls from the quality of interaction experience. Still, the design does not permit causal claims; the results should be read as strong associations consistent with the hypothesis, not as definitive proof of causation.

7.3. Methodological Considerations

Several analysis choices were made to keep inference aligned with the design and to guard against small-sample pitfalls. For H1, the Main phase participant mean was chosen as the primary estimand because that is where both groups operated under the same control and task rules. The mixed model at the run level, supplemented by label-permutation inference when random-effects estimates became degenerate, allowed the analysis to acknowledge dependence without relying on asymptotic Wald tests that are fragile at $N = 28$. The ANCOVA served a specific purpose, showing the Drag advantage is not an artefact of Intervention baseline, and the non-significant $group \times baseline$ interaction provided a direct test of the homogeneity-of-slopes assumption often left implicit.

For H2, using robust standard errors in OLS guarded against small-sample variance difference on the $[-3, 3]$ UEQ scales. Using ordinal logistic and Spearman for MCQ5/MCQ6 respected their ordered-category nature while removing the ”6 = not applicable” category from the intensity scale. Because the study involves families of parallel tests (6 scales, 26 items), the FDR was controlled rather than the family-wise error rate. This choice is appropriate when the goal is to identify a pattern of effects without becoming so conservative that genuine signals are erased, especially at this sample size. Finally, the distinction between objective and perceived interaction intensity was formalised in the models via time adjustments and by placing the perceptual metric as the primary test for the second research question. This is faithful to the question’s wording about ”user-driven interaction intensity” and ”perceived engagement”.

7.4. Limitations and Alternative Readings

The most obvious limitation is statistical power. With 28 participants, the study is well-suited to detect moderate to large effects, but will produce borderline p-values for moderate effects in the $0.5 - 0.7$ SD range, as observed for the first research question. That does not negate the practical magnitude of the differences, but it calls for caution against over-generalisation. The second limitation is zero inflation in objective interaction with many participants gathering little or no time on reactive elements; linear models on log time only partially address skew. A two-part model could be considered in future work to separate the decision to interact at all from the amount of interaction among interactors. A third limitation is measurement scope. Objective time collapses ”what was done” into ”how long”. It misses important distinctions such as purposeful actions versus incidental contact, temporal alignment of touches with goals, and the tactile extent of events. Such features could sharpen behavioural links to user experience and might reveal effects that total time alone cannot.

The following design and measurement constraints could be read as alternative explanations for some effects. The onboarding aimed for minimal guidance to preserve naturalistic exploration. While this choice fits the study’s spirit, a more structured tutorial that potentially includes an intuition-building base on hyperbolic geometry might have reduced early variance and cold-start costs in the Main task. The trade-off is that a stronger structuring could reduce novelty effects, but risks teaching to the test and diluting practical validity.

There is also a training-evaluation mismatch. During the Intervention, participants practised on a 2D plane (Minimap); in Main, they gained an extra axis through self-rotation while walking that can be off-putting. Several participants visibly adapted by fixing a local orientation and advancing with side/forward steps, while others rotated freely. This strategy discrepancy was not modelled explicitly and may have contributed to between-participant variance that is not attributed to the control scheme they used. Relatedly, first-exposure effects are likely; most participants had no prior experience with a space of this kind, so part of what was measured is the cost of learning to predict a novel control-geometry combination.

Additionally, two task-level constraints limit generality. First, the ”optimal path” used for the Efficiency and Sequence pillars is a principled alternative, but under Holonomy VR conditions, the notion of optimality depends on local frames. This ground truth may slightly over- or under-penalise certain routes. Second, the study used two levels per phase to keep the sessions tractable and of reasonable duration. However, this restricts the spread of difficulties and the precision of run-level variance estimates. Additional, diverse levels would better separate learning, difficulty, and interface effects, and would allow stronger tests of path optimality under explicit Holonomy VR constraints.

It is also noted that the UEQ and MCQ6 have a post-treatment nature. Both are outcomes of the experience and susceptible to shared method variance. The time-adjusted analyses partly address the simplest confound, but causal claims require a design that manipulates interaction intensity directly. Finally, the results are task-specific. The performance composite is tuned to non-Euclidean navigation under this environment and these haptic responses; different tasks or reactive element designs may alter which UEQ facets respond most strongly.

7.5. Implications

The results regarding the first research question imply that once participants are placed on equal footing in Main, those who had trained with Drag carry over a performance advantage. In practical terms, Drag's platform facilitates embodiment by associating hand/arm movements with movements on the minimap/world. This continuous mapping may support a more efficient transformation in non-Euclidean layouts, yielding better timing, path choice, and action sequencing. This bears on interface design, when the environment departs from familiar Euclidean structure, control schemes that reduce translation costs can pay off at evaluation time.

The second research question's results point to the importance of how reactive interactions are experienced, not merely how long they occur. The strong, time-robust link between MCQ6 and UEQ Efficiency suggests that engaging, memorable contact with reactive elements supports pragmatic usability. The supporting signals on Dependability and Stimulation fit the story; when interactions themselves feel "right", the environment feels more predictable and exciting. For embodied learning accounts, this pattern of: objective time weak, perceived engagement strong, suggests that the quality and meaning of contact, not mere duration, is central. This aligns with enactive perspectives; it is the sensorimotor experience achieved that predicts perceived efficiency, not just clocked exposure.

7.6. Positioning within the State of the Art

This thesis sits at the overlap of three conversations that are often kept apart: control and performance in unfamiliar (non-Euclidean) environments, embodied accounts of how action shapes perception, and user-experience work on the role of reactive haptic feedback in VR that is not strictly task-critical. Bringing them together in one task, with one set of measures, allows for saying something concrete about *when* a continuous, embodied interaction helps and *how* voluntary, tactile exploration shows up in people's experience.

On the control side, most prior work on navigation in non-Euclidean spaces has shown that people do not merely transpose Euclidean intuitions; instead, they learn local action-outcome regularities, often with increased cognitive load and unfamiliar path structure. Much of that literature has focused on spatial cognition, how well participants can form and use internal representations, under visual and movement variants that keep the control method relatively simple. The present results extend that line by touching upon the control method itself. When the mapping is continuous (Drag) rather than discretised (Buttons), the Main-phase performance is consistently higher in equal conditions, even once baseline performance is accounted for. In other words, when the geometry is already asking the user to update their predictive model of movement, a control scheme that minimises translation between intention and act appears to reduce error across timing, path choice, and action ordering. This is precisely where an embodied perspective on control predicts an advantage: fewer "transformations" in the control loop, tighter perception-action coupling, less overhead devoted to managing the interface itself. Methodologically, this is tied to the design rather than a single estimator. Participant-mean comparisons, a run-level mixed model tested by participant-label permutation, and ANCOVA against Intervention all point the same way.

On the experience side, reactive haptics are often discussed as "nice to have" for presence, realism, or delight, with mixed evidence about later impacts on task performance or usability. Here, more specificity is afforded. A question is whether "non-essential" haptics (touching bushes and trees in this case) meaningfully shape experience or simply decorate it. The pattern observed in this research clarifies this distinction. Simply spending more seconds touching reactive elements is a weak predictor of UEQ scales. What matters is whether those touches feel engaging and memorable. The MCQ6→UEQ link, even after adjusting for exposure and objective time, suggests that meaningful interactions at the sensorimotor level are the active ingredient. Item-level results also reinforce this. In short, hedonic engagement and pragmatic usability are not orthogonal in this setting. When the interaction ”clicks”, the system also feels more efficient. The contribution here is not a claim that ”more haptics is always better”, but a more precise and testable proposition that the quality of sensorimotor coupling and not just duration, adapts with actual usability judgments.

A second contribution is measurement clarity. Exposure time, behaviour time, and experience are deliberately separated and then tested for links among them. Exposure and behaviour are related but zero-inflated; behaviour and experience are related but only moderately so. The strongest, most reliable association with UEQ runs through experience. This helps explain why ”more haptics” or ”more time on task” sometimes fail to move UX. Exposure is necessary, but engagement is what translates into perceived efficiency. This also rationalises the mixed results around ”training transfer” in complex VR tasks. Improvements may depend less on how long users touch reactive elements and more on whether those contacts afford learning-relevant predictors that later shape action selection. By making these distinctions explicit and showing their different outcomes, this research offers a useful template for future VR UX studies to avoid over-attributing effects to time alone.

Finally, the analysis choices are part of the contribution. Estimates were matched to the design (Main-phase as primary target), dependence is handled explicitly (mixed models), used permutations when random-effect estimates collapsed, relied on robust errors and rank checks for small-sample inference, and multiplicity is controlled with FDR at the scale and item families. That combination can serve as a template and platform for similar future studies.

Against this background, this thesis advances the state of the art in two ways. First, it provides evidence that continuous, embodied control can yield a practically meaningful performance advantage in a non-Euclidean environment under uniform operating conditions. This answers the first research question to the degree allowed by the sample; not as definitive proof, but as a robust effect that future, larger studies can work on. Second, it shows that perceived engagement with non-essential reactive elements is reliably associated with perceived efficiency, even when adjusting for interaction time. That answers the second research question at its core; it is not the existence of reactive elements or the raw time spent with them that matter most, but whether those interactions form the experience that users find engaging, and in turn more efficient, stimulating, and useful. In both cases, the contribution is progressive rather than final. The work set an experimental and analytical platform that others can reuse to move from association to causation and from a single task to broader generality.

8 Conclusion

This thesis examined how embodied control and voluntary interaction shaped both performance and experience in a non-Euclidean, haptically enabled VR environment. Two questions guided the study. RQ1 asked whether participants trained with a continuous, drag-based control would later perform better than those trained with button-based input when evaluated under identical conditions. RQ2 asked how naturally occurring variation in interaction with reactive elements (trees, hedges) relates to user experience during the main task.

Across the analyses that matter for the design, participant-level comparisons in the Main phase, a dependence-aware mixed model with permutation inference, and an ANCOVA that adjusts for Intervention, Drag consistently outperformed Buttons by a practically meaningful margin. The estimates were stable across methods and pointed in the same direction, while acknowledging the statistical modesty expected at this sample size. The take-home message is straightforward: when both groups face the same non-Euclidean task, prior practice with a continuous mapping carries over to better overall fluency (speed, path, sequencing), which is exactly what an embodied account of control would predict.

The clearest way from ”interaction intensity” to user experience ran through perceived engagement/ memorability (MCQ6), not through raw time. Participants who experienced the reactive contacts as engaging reported higher UEQ scores. This association persisted even after controlling for measured exposure. Item-level results reinforced the picture by showing which anchors of the UEQ moved with engagement. By contrast, objective interaction time and self-estimated time were weak predictors of UEQ in this dataset. Together, these findings suggest that in non-Euclidean VR, it is the quality of the tactile exploration, rather than its duration, that most reliably translates into better reported usability.

Beyond specific answers, the work contributes an analysis pattern for small-N VR: a transparent pillar-based performance score, estimands aligned to the design, with mixed models, and multiplicity-aware robust scale/item analyses for UEQ. Separating exposure, behaviour, and experience as distinct constructs proved especially useful for avoiding common pitfalls in the literature and for explaining why ”more time” does not automatically produce better UX.

There are inevitably limits. With 28 participants, the precision around moderate effects is finite. Objective interaction time is zero-inflated and unrefined as a behavioural summary. Both MCQ6 and UEQ are post-treatment self-reports, so the experience findings are associative, not causal. These constraints do not undercut the pattern but do set the scope: the claims are strongest for the task studied and for the way reactive elements were implemented here.

The results suggest concrete next steps. Increasing the sample size would tighten uncertainty around the Main-phase group difference. Manipulating the quality and intensity of reactive interactions would convert the engagement-UX link from association to causal estimate. Richer telemetry would clarify how different aspects of exploration matter. Cross-over study designs could test the durability and transfer of embodied training benefits.

In closing, the study offers two simple lessons for designing and evaluating interaction in less conventional VR spaces. First, minimise translation costs in the control loop; continuous, embodied mappings better support later proficiency when the world itself is unfamiliar. Second, design a reactive contact that invites meaningful exploration; it is the felt engagement that users turn into a sense of efficiency and control. Put differently, in spaces where straight lines bend, how we move and how we choose to touch the world are what make it navigable.

Bibliography

[1] Dor Abrahamson and Raúl Sánchez-García. “Learning Is Moving in New Ways: The Ecological Dynamics of Mathematics Education”. In: *Journal of the Learning Sciences* 25.2 (Apr. 2, 2016). Publisher: Routledge _eprint: https://doi.org/10.1080/10508406.2016.1143370, pp. 203–239. ISSN: 1050-8406. DOI: 10.1080/10508406.2016.1143370. URL: https://doi.org/10.1080/10508406.2016.1143370 (visited on 03/31/2025).

[2] Hamza Afzaal and Usman Alim. *Haptic Stylus vs. Handheld Controllers: A Comparative Study for Surface Visualization Interactions*. Dec. 10, 2024. DOI: 10.48550/arXiv.2412.07065. arXiv: 2412.07065[cs]. URL: http://arxiv.org/abs/2412.07065 (visited on 01/10/2025).

[3] Martha W. Alibali and Mitchell J. Nathan. “Embodiment in Mathematics Teaching and Learning: Evidence From Learners’ and Teachers’ Gestures”. In: *Journal of the Learning Sciences* 21.2 (Apr. 2012), pp. 247–286. ISSN: 1050-8406, 1532-7809. DOI: 10.1080/10508406.2011.611446. URL: http://www.tandfonline.com/doi/abs/10.1080/10508406.2011.611446 (visited on 05/28/2025).

[4] Rena Bayramova et al. “The role of vision and proprioception in self-motion encoding: An immersive virtual reality study”. In: *Attention, Perception & Psychophysics* 83.7 (Oct. 2021), pp. 2865–2878. ISSN: 1943-393X. DOI: 10.3758/s13414-021-02344-8.

[5] Kenan Bektaş et al. “The systematic evaluation of an embodied control interface for virtual reality”. In: *PLoS ONE* 16.12 (Dec. 7, 2021), e0259977. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0259977. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8651145/ (visited on 06/04/2025).

[6] Lewy Blue (looeee). *Order 5 square tiling on Poincaré Disk*. GitHub. Apr. 2, 2021. URL: https://raw.githubusercontent.com/looeee/hyperbolic-tiling/main/assets/uniform-hyperbolic-tiling-45.png (visited on 06/18/2025).

[7] Manuel Caeiro-Rodríguez et al. “A Systematic Review of Commercial Smart Gloves: Current Status and Applications”. In: *Sensors* 21.8 (Jan. 2021). Number: 8 Publisher: Multidisciplinary Digital Publishing Institute, p. 2667. ISSN: 1424-8220. DOI: 10.3390/s21082667. URL: https://www.mdpi.com/1424-8220/21/8/2667 (visited on 05/29/2025).

[8] Julia Chatain, Manu Kapur, and Robert W. Sumner. “Three Perspectives on Embodied Learning in Virtual Reality: Opportunities for Interaction Design”. In: *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI EA ’23. New York, NY, USA: Association for Computing Machinery, Apr. 19, 2023, pp. 1–8. ISBN: 978-1-4503-9422-2. DOI: 10.1145/3544549.3585805. URL: https://dl.acm.org/doi/10.1145/3544549.3585805 (visited on 02/03/2025).

[9] Umer Asghar Chattha et al. “Motion Sickness in Virtual Reality: An Empirical Evaluation”. In: *IEEE Access* 8 (2020), pp. 130486–130499. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.3007076. URL: https://ieeexplore.ieee.org/abstract/document/9133071 (visited on 06/26/2025).

[10] Lucia Cherep et al. *Spatial cognitive implications of teleporting through virtual environments*. May 31, 2019. DOI: 10.31234/osf.io/cx9vt. URL: https://osf.io/cx9vt_v1 (visited on 06/04/2025).

[11] Chris G. Christou and Poppy Aristidou. “Steering Versus Teleport Locomotion for Head Mounted Displays”. In: *Augmented Reality, Virtual Reality, and Computer Graphics*. International Conference on Augmented Reality, Virtual Reality and Computer Graphics. ISSN: 1611-3349. Springer, Cham, 2017, pp. 431–446. ISBN: 978-3-319-60928-7. DOI: 10.1007/978-3-319-60928-7_37. URL: https://link.springer.com/chapter/10.1007/978-3-319-60928-7_37 (visited on 06/08/2025).

[12] Alessandro Forgiarini, Fabio Buttussi, and Luca Chittaro. “Virtual Reality for Object Location Spatial Memory: A Comparison of Handheld Controllers and Force Feedback Gloves”. In: *Proceedings of the 15th Biannual Conference of the Italian SIGCHI Chapter*. CHItaly 2023: 15th Biannual Conference of the Italian SIGCHI Chapter. Torino Italy: ACM, Sept. 20, 2023, pp. 1–9. ISBN: 979-8-4007-0806-0. DOI: 10.1145/3605390.3605403. URL: https://dl.acm.org/doi/10.1145/3605390.3605403 (visited on 03/31/2025).

[13] Doron Friedman et al. “Navigating Virtual Reality by Thought: What Is It Like?” In: *Presence: Teleoperators and Virtual Environments* 16.1 (Feb. 1, 2007), pp. 100–110. ISSN: 1054-7460. DOI: 10.1162/pres.16.1.100. URL: https://direct.mit.edu/pvar/article/16/1/100-110/18658 (visited on 05/29/2025).

[14] BoYu Gao et al. “Exploring Bimanual Haptic Feedback for Spatial Search in Virtual Reality”. In: *IEEE Transactions on Visualization and Computer Graphics* 30.5 (May 2024). Conference Name: IEEE Transactions on Visualization and Computer Graphics, pp. 2422–2433. ISSN: 1941-0506. DOI: 10.1109/TVCG.2024.3372045. URL: https://ieeexplore.ieee.org/document/10458369 (visited on 02/03/2025).

[15] Janet K. Gibbs, Marco Gillies, and Xueni Pan. “A comparison of the effects of haptic and visual feedback on presence in virtual reality”. In: *International Journal of Human-Computer Studies* 157 (Jan. 1, 2022), p. 102717. ISSN: 1071-5819. DOI: 10.1016/j.ijhcs.2021.102717. URL: https://www.sciencedirect.com/science/article/pii/S107158192100135X (visited on 03/31/2025).

[16] User Harkonnen2. *Diagram illustrating Euclid’s parallel postulate*. Wikipedia. Aug. 14, 2008. URL: https://commons.wikimedia.org/wiki/File:Parallel_postulate_en.svg (visited on 07/02/2025).

[17] Andreas Hinderks, Martin Schrepp, and Jörg Thomaschewski. *(PDF) User Experience Questionnaire Handbook*. ResearchGate. DOI: 10.13140/RG.2.1.2815.0245. URL: https://www.researchgate.net/publication/281973617_User_Experience_Questionnaire_Handbook (visited on 04/10/2025).

[18] Autumn B. Hostetter and Martha W. Alibali. “Visible embodiment: Gestures as simulated action”. In: *Psychonomic Bulletin & Review* 15.3 (June 1, 2008). Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 3 Publisher: Springer-Verlag, pp. 495–514. ISSN: 1531-5320. DOI: 10.3758/PBR.15.3.495. URL: https://link.springer.com/article/10.3758/PBR.15.3.495 (visited on 05/28/2025).

[19] Helen H. Huang, Hanspeter Pfister, and Yalong Yang. “Is Embodied Interaction Beneficial? A Study on Navigating Network Visualizations”. In: *Information Visualization* 22.3 (July 2023), pp. 169–185. ISSN: 1473-8716, 1473-8724. DOI: 10.1177/14738716231157082. arXiv: 2301.11516[cs]. URL: http://arxiv.org/abs/2301.11516 (visited on 06/04/2025).

[20] *Hyperbolica*. Hyperbolica. Mar. 14, 2022. URL: https://store.steampowered.com/app/1256230/Hyperbolica/ (visited on 04/01/2025).

[21] *HyperRogue*. HyperRogue. Jan. 16, 2015. URL: https://store.steampowered.com/app/342610/HyperRogue/ (visited on 06/08/2025).

[22] Eloy Irigoyen, Mikel Larrea, and Manuel Graña. “A Narrative Review of Haptic Technologies and Their Value for Training, Rehabilitation, and the Education of Persons with Special Needs”. In: *Sensors* 24.21 (Jan. 2024). Number: 21 Publisher: Multidisciplinary Digital Publishing Institute, p. 6946. ISSN: 1424-8220. DOI: 10.3390/s24216946. URL: https://www.mdpi.com/1424-8220/24/21/6946 (visited on 05/29/2025).

[23] Alexander Jaksties, Achim Ebert, and Kerstin Müller. “Magical Tower VR - Examining Orientation in Non-Euclidean Spaces”. In: *Proceedings of the 33rd European Conference on Cognitive Ergonomics*. ECCE ’22. New York, NY, USA: Association for Computing Machinery, Oct. 4, 2022, pp. 1–6. ISBN: 978-1-4503-9808-4. DOI: 10.1145/3552327.3552344. URL: https://dl.acm.org/doi/10.1145/3552327.3552344 (visited on 06/04/2025).

[24] Mina C. Johnson-Glenberg et al. “Collaborative embodied learning in mixed reality motion-capture environments: Two science studies.” In: *Journal of Educational Psychology* 106.1 (2014), pp. 86–104. ISSN: 1939-2176, 0022-0663. DOI: 10.1037/a0034008. URL: https://doi.apa.org/doi/10.1037/a0034008 (visited on 05/27/2025).

[25] Kazushige Kimura et al. “Orientation in Virtual Reality Does Not Fully Measure Up to the Real-World”. In: *Scientific Reports* 7.1 (Dec. 22, 2017). Publisher: Nature Publishing Group, p. 18109. ISSN: 2045-2322. DOI: 10.1038/s41598-017-18289-8. URL: https://www.nature.com/articles/s41598-017-18289-8 (visited on 04/02/2025).

[26] Sabine U. König et al. “Embodied Spatial Knowledge Acquisition in Immersive Virtual Reality: Comparison to Map Exploration”. In: *Frontiers in Virtual Reality* 2 (Mar. 3, 2021). Publisher: Frontiers. ISSN: 2673-4192. DOI: 10.3389/frvir.2021.625548. URL: https://www.frontiersin.org/journals/virtual-reality/articles/10.3389/frvir.2021.625548/full (visited on 03/31/2025).

[27] Eike Langbehn, Paul Lubos, and Frank Steinicke. “Evaluation of Locomotion Techniques for Room-Scale VR: Joystick, Teleportation, and Redirected Walking”. In: *Proceedings of the Virtual Reality International Conference - Laval Virtual*. VRIC ’18. New York, NY, USA: Association for Computing Machinery, Apr. 4, 2018, pp. 1–9. ISBN: 978-1-4503-5381-6. DOI: 10.1145/3234253.3234291. URL: https://dl.acm.org/doi/10.1145/3234253.3234291 (visited on 06/08/2025).

[28] Robb Lindgren and Mina Johnson-Glenberg. “Emboldened by Embodiment: Six Precepts for Research on Embodied Learning and Mixed Reality”. In: *Educational Researcher* 42.8 (2013). Publisher: [American Educational Research Association, Sage Publications, Inc.], pp. 445–452. ISSN: 0013-189X. URL: https://www.jstor.org/stable/24571228 (visited on 05/27/2025).

[29] O. A. J. van der Meijden and M. P. Schijven. “The value of haptic feedback in conventional and robot-assisted minimal invasive surgery and virtual reality training: a current review”. In: *Surgical Endoscopy* 23.6 (June 2009), pp. 1180–1190. ISSN: 0930-2794. DOI: 10.1007/s00464-008-0298-x. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2686803/ (visited on 05/26/2025).

[30] Thinh Nguyen-Vo et al. “NaviBoard and NaviChair: Limited Translation Combined with Full Rotation for Efficient Virtual Locomotion”. In: *IEEE Transactions on Visualization and Computer Graphics* 27.1 (Jan. 2021). Conference Name: IEEE Transactions on Visualization and Computer Graphics, pp. 165–177. ISSN: 1941-0506. DOI: 10.1109/TVCG.2019.2935730. URL: https://ieeexplore.ieee.org/abstract/document/8809840 (visited on 04/02/2025).

[31] Raphael Palombo et al. “Glove versus controller: the effect of VR gloves and controllers on presence, embodiment, and cognitive absorption”. In: *Frontiers in Virtual Reality* 5 (Mar. 26, 2024). Publisher: Frontiers. ISSN: 2673-4192. DOI: 10.3389/frvir.2024.1337959. URL: https://www.frontiersin.org/journals/virtual-reality/articles/10.3389/frvir.2024.1337959/full (visited on 03/31/2025).

[32] Randy Pausch, Dennis Proffitt, and George Williams. “Quantifying immersion in virtual reality”. In: *Proceedings of the 24th annual conference on Computer graphics and interactive techniques - SIGGRAPH ’97*. the 24th annual conference. Not Known: ACM Press, 1997, pp. 13–18. ISBN: 978-0-89791-896-1. DOI: 10.1145/258734.258744. URL: http://portal.acm.org/citation.cfm?doid=258734.258744 (visited on 01/10/2025).

[33] Vincent Adelizzi Pisani et al. “Navigation by Walking In Hyperbolic Space Using Virtual Reality”. In: *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. CHI PLAY ’19 Extended Abstracts. New York, NY, USA: Association for Computing Machinery, Oct. 17, 2019, pp. 611–618. ISBN: 978-1-4503-6871-1. DOI: 10.1145/3341215.3356287. URL: https://doi.org/10.1145/3341215.3356287 (visited on 01/17/2025).

[34] Emmanuel B. Vander Poorten et al. *To Feedback or not to Feedback - the Value of Haptics in Virtual Reality Surgical Training*. Artwork Size: 5 pages ISBN: 9783905674767 Pages: 5 pages Publication Title: EuroVR 2014 - Conference and Exhibition of the European Association of Virtual and Augmented Reality. 2014. DOI: 10.2312/EUROVR.20141356. URL: http://diglib.eg.org/handle/10.2312/eurovr.20141356.157-161 (visited on 05/29/2025).

[35] Antony Prakash and Ramkumar Rajendran. “Virtual Reality and Embodied Learning: Unraveling the Relationship via Dynamic Learner Behavior”. In: *International Conference on Computers in Education* (Dec. 4, 2023). ISSN: 3078-4360. DOI: 10.58459/icce.2023.1412. URL: https://library.apsce.net/index.php/ICCE/article/view/1412 (visited on 06/04/2025).

[36] Adam S Radomsky et al. “The Claustrophobia Questionnaire”. In: *Journal of Anxiety Disorders* 15.4 (July 2001). Publisher: Elsevier BV, pp. 287–297. ISSN: 0887-6185. DOI: 10.1016/s0887-6185(01)00064-0. URL: https://linkinghub.elsevier.com/retrieve/pii/S0887618501000640 (visited on 07/08/2025).

[37] Hedvig Reuterswärd. “Mission Climbossible - a study of immersive vertical locomotion in impossible spaces for virtual reality”. In: (2020). URL: https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-280844.

[38] Roy A. Ruddle and Simon Lessels. “For efficient navigational search, humans require full physical movement, but not a rich visual scene”. In: *Psychological Science* 17.6 (June 2006), pp. 460–465. ISSN: 0956-7976. DOI: 10.1111/j.1467-9280.2006.01728.x.

[39] Roy A. Ruddle, Ekaterina Volkova, and Heinrich H. Bülthoff. “Walking improves your cognitive map in environments that are large-scale and large in extent”. In: *ACM Trans. Comput.-Hum. Interact.* 18.2 (July 1, 2011), 10:1–10:20. ISSN: 1073-0516. DOI: 10.1145/1970378.1970384. URL: https://dl.acm.org/doi/10.1145/1970378.1970384 (visited on 05/29/2025).

[40] Greg S. Ruthenbeck, Michael Tlauka, and Andria Tan. “Does Haptic Feedback Improve Learning and Recall of Spatial Information? A Study Using a Virtual Reality Nasendoscopy Simulation”. In: *Haptics: Perception, Devices, Control, and Applications*. International Conference on Human Haptic Sensing and Touch Enabled Computer Applications. ISSN: 1611-3349. Springer, Cham, 2016, pp. 404–411. ISBN: 978-3-319-42321-0. DOI: 10.1007/978-3-319-42321-0_37. URL: https://link.springer.com/chapter/10.1007/978-3-319-42321-0_37 (visited on 02/03/2025).

[41] *SenseGlove Nova 1 — SenseGlove Docs documentation*. URL: https://senseglove.gitlab.io/SenseGloveDocs/nova-glove.html (visited on 04/01/2025).

[42] Yuxiang Shi and Guozhen Shen. “Haptic Sensing and Feedback Techniques toward Virtual Reality”. In: *Research* 7 (2024), p. 0333. ISSN: 2639-5274. DOI: 10.34133/research.0333. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10964227/ (visited on 02/03/2025).

[43] Martin Skrodzki. *Illustrations of non-Euclidean geometry in virtual reality*. Sept. 10, 2020. DOI: 10.48550/arXiv.2008.01363. arXiv: 2008.01363[math]. URL: http://arxiv.org/abs/2008.01363 (visited on 01/10/2025).

[44] Martin Skrodzki et al. “Holonomy: A Virtual Reality Exploration of Hyperbolic Geometry”. In: *Proceedings of the 29th International ACM Conference on 3D Web Technology*. WEB3D ’24: The 29th International ACM Conference on 3D Web Technology. Guimarães Portugal: ACM, Sept. 25, 2024, pp. 1–10. ISBN: 979-8-4007-0689-9. DOI: 10.1145/3665318.3677149. URL: https://dl.acm.org/doi/10.1145/3665318.3677149 (visited on 01/10/2025).

[45] Martin Usoh et al. “Walking > walking-in-place > flying, in virtual environments”. In: *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. SIGGRAPH ’99. USA: ACM Press/Addison-Wesley Publishing Co., July 1, 1999, pp. 359–364. ISBN: 978-0-201-48560-8. DOI: 10.1145/311535.311589. URL: https://dl.acm.org/doi/10.1145/311535.311589 (visited on 06/03/2025).

[46] A. de Vries. “Mapping hyperbolic space for the virtual reality game ”Holonomy””. In: (2023). URL: https://repository.tudelft.nl/record/uuid:b86007e3-cd1a-4654-8556-09430bf75a99 (visited on 06/04/2025).

[47] David Waller and Eric Hodgson. “Sensory Contributions to Spatial Knowledge of Real and Virtual Environments”. In: *Human Walking in Virtual Environments*. Springer, New York, NY, 2013, pp. 3–26. ISBN: 978-1-4419-8432-6. DOI: 10.1007/978-1-4419-8432-6_1. URL: https://link.springer.com/chapter/10.1007/978-1-4419-8432-6_1 (visited on 05/29/2025).

[48] Christopher Widdowson and Ranxiao Frances Wang. “Human navigation in curved spaces”. In: *Cognition* 218 (Jan. 1, 2022), p. 104923. ISSN: 0010-0277. DOI: 10.1016/j.cognition.2021.104923. URL: https://www.sciencedirect.com/science/article/pii/S0010027721003462 (visited on 03/31/2025).

[49] Christopher Widdowson and Ranxiao Frances Wang. “Human spatial learning strategies in wormhole virtual environments”. In: *Spatial Cognition & Computation* 23.2 (Apr. 3, 2023), pp. 112–131. ISSN: 1387-5868, 1542-7633. DOI: 10.1080/13875868.2022.2095277. URL: https://www.tandfonline.com/doi/full/10.1080/13875868.2022.2095277 (visited on 03/31/2025).

[50] Andrew D. Wilson and Sabrina Golonka. “Embodied Cognition is Not What you Think it is”. In: *Frontiers in Psychology* 4 (Feb. 12, 2013). Publisher: Frontiers. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2013.00058. URL: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2013.00058/full (visited on 03/31/2025).

A

Extra results

Scale	Mean	SD	Confidence	95%CI
Attractiveness	1.440	1.131	0.419	$[1.021, 1.860]$
Perspicuity	0.571	1.388	0.514	$[0.057, 1.085]$
Efficiency	1.071	1.182	0.438	$[0.634, 1.509]$
Dependability	0.741	1.044	0.387	$[0.354, 1.128]$
Stimulation	1.438	0.925	0.342	$[1.095, 1.780]$
Novelty	1.813	0.841	0.311	$[1.501, 2.124]$

Table A.1: UEQ scale means and descriptives from all participants for the Intervention phase.

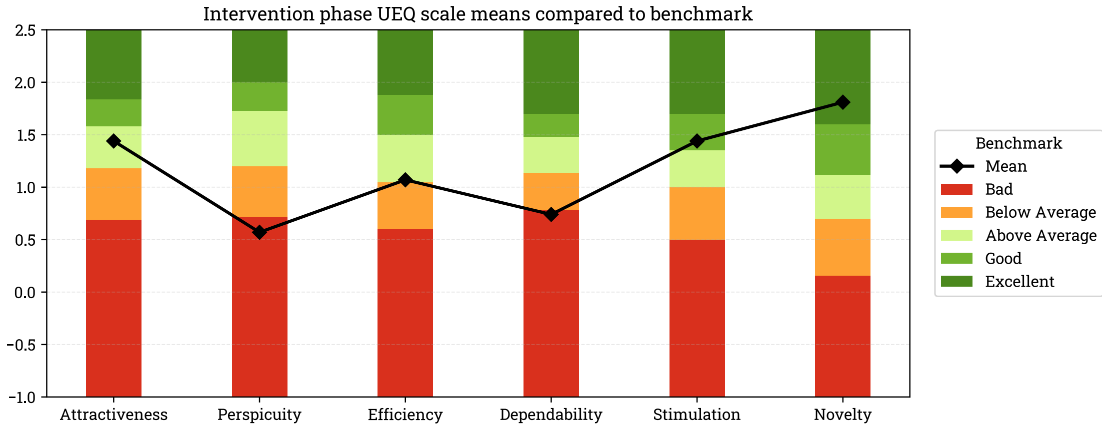


Figure A.1: The UEQ scale means from the Intervention phase responses compared to a benchmark.

UEQ Scale	β	95%CI	p-value	FDR q	Spearman's ρ	Spearman's p-value
Attractiveness	0.554	$[-0.252, 1.361]$	0.178	0.208	0.409	0.031
Perspicuity	0.518	$[-0.200, 1.236]$	0.157	0.208	0.275	0.157
Efficiency	0.811	$[0.288, 1.335]$	0.002	0.014	0.627	0.000
Dependability	0.552	$[-0.054, 1.157]$	0.074	0.148	0.326	0.091
Stimulation	0.420	$[-0.036, 0.875]$	0.071	0.148	0.359	0.061
Novelty	0.420	$[-0.110, 0.506]$	0.208	0.208	0.268	0.169

Table A.2: Results of OLS regression model for each UEQ scale on $zMCQ6$, adjusted for time with zI_{total} and zT_{main}. Spearman's ρ and FDR control included.

UEQ Scale	β	95%CI	p-value	FDR q	Spearman's ρ	Spearman's p-value
Attractiveness	-0.033	$[-0.317, 0.251]$	0.821	0.833	0.002	0.994
Perspicuity	-0.225	$[-0.685, 0.234]$	0.337	0.685	-0.168	0.413
Efficiency	-0.161	$[-0.493, 0.171]$	0.342	0.685	-0.169	0.408
Dependability	-0.086	$[-0.433, 0.261]$	0.628	0.833	-0.078	0.705
Stimulation	-0.124	$[-0.320, 0.073]$	0.217	0.685	-0.227	0.266
Novelty	0.027	$[-0.227, 0.281]$	0.833	0.833	-0.004	0.986

Table A.3: Results of OLS regression model for each UEQ scale on $zMCQ5$, with Spearman's ρ and FDR control included.

UEQ Scale	β	95%CI	p-value	FDR q	Spearman's ρ	Spearman's p-value
Attractiveness	-0.084	$[-0.502, 0.335]$	0.696	0.915	0.002	0.994
Perspicuity	-0.193	$[-0.823, 0.438]$	0.549	0.915	-0.168	0.413
Efficiency	-0.206	$[-0.741, 0.328]$	0.449	0.915	-0.169	0.408
Dependability	-0.023	$[-0.510, 0.464]$	0.925	0.925	-0.078	0.705
Stimulation	-0.093	$[-0.510, 0.464]$	0.557	0.915	-0.227	0.266
Novelty	0.060	$[-0.331, 0.452]$	0.762	0.915	-0.004	0.986

Table A.4: Results of OLS regression model for each UEQ scale on $zMCQ5$, adjusted for time with zI_{total} and zT_{main}. Spearman's ρ and FDR control included.

UEQ Scale	β	95%CI	p-value	FDR q	Spearman's ρ	Spearman's p-value
Attractiveness	0.104	$[-0.399, 0.606]$	0.686	0.944	0.205	0.294
Perspicuity	-0.117	$[-0.589, 0.355]$	0.627	0.944	-0.142	0.471
Efficiency	0.118	$[-0.344, 0.579]$	0.617	0.944	0.066	0.738
Dependability	-0.018	$[-0.514, 0.478]$	0.944	0.944	-0.068	0.732
Stimulation	-0.013	$[-0.377, 0.351]$	0.944	0.944	-0.068	0.732
Novelty	0.080	$[-0.162, 0.322]$	0.518	0.944	0.106	0.592

Table A.5: Results of OLS regression model for each UEQ scale on zI_{total}, with Spearman's ρ and FDR control included.

UEQ Scale	β	95%CI	p-value	FDR q	Spearman's ρ	Spearman's p-value
Attractiveness	0.092	[-0.532, 0.715]	0.774	0.945	0.205	0.294
Perspicuity	-0.103	[-0.633, 0.427]	0.703	0.945	-0.142	0.471
Efficiency	0.151	[-0.423, 0.725]	0.606	0.945	0.066	0.738
Dependability	0.021	[-0.571, 0.612]	0.945	0.945	-0.068	0.732
Stimulation	0.029	[-0.399, 0.457]	0.894	0.945	-0.068	0.732
Novelty	0.114	[-0.167, 0.395]	0.427	0.945	0.106	0.592

Table A.6: Results of OLS regression model for each UEQ scale on zI_{total}, adjusted for time with zT_{main}. Spearman's ρ and FDR control included.

UEQ Scale	Item #	β	95%CI	p-value	FDR q	Spearman's ρ	Spearman's p-value
Attractiveness	16	0.858	[0.186, 1.529]	0.012	0.054	0.522	0.004
	14	0.613	[−0.084, 1.311]	0.085	0.220	0.349	0.069
	12	0.508	[−0.111, 1.128]	0.108	0.255	0.270	0.164
	1	0.563	[−0.359, 1.484]	0.231	0.310	0.288	0.138
	25	0.350	[−0.232, 0.933]	0.238	0.310	0.306	0.113
	24	0.337	[−0.307, 0.981]	0.305	0.367	0.204	0.298
Perspicuity	13	0.588	[−0.036, 1.212]	0.065	0.187	0.344	0.073
	2	0.522	[−0.133, 1.176]	0.118	0.256	0.247	0.205
	21	0.372	[−0.348, 1.092]	0.311	0.367	0.227	0.245
	4	0.261	[−0.485, 1.007]	0.493	0.558	0.073	0.713
Efficiency	20	0.912	[0.468, 1.356]	0.000	0.001	0.550	0.002
	23	1.013	[0.448, 1.578]	0.000	0.006	0.495	0.007
	22	0.757	[0.170, 1.344]	0.012	0.054	0.450	0.016
	9	0.381	[−0.234, 0.996]	0.224	0.310	0.221	0.258
Dependability	11	0.654	[0.173, 1.135]	0.008	0.050	0.391	0.040
	17	0.563	[0.006, 1.119]	0.048	0.155	0.414	0.028
	19	0.483	[−0.283, 1.249]	0.216	0.310	0.215	0.272
	8	0.256	[−0.565, 1.078]	0.541	0.586	0.024	0.905
Stimulation	5	0.634	[0.119, 1.150]	0.016	0.059	0.448	0.017
	18	0.336	[−0.112, 0.784]	0.142	0.265	0.225	0.249
	7	0.273	[−0.152, 0.698]	0.208	0.310	0.249	0.201
	6	0.234	[−0.132, 0.601]	0.210	0.310	0.321	0.096
Novelty	26	0.509	[0.173, 1.135]	0.002	0.014	0.478	0.010
	10	0.244	[−0.082, 0.571]	0.143	0.265	0.300	0.120
	3	0.070	[0.070, −0.267]	0.685	0.712	0.115	0.560
	15	−0.042	[−0.516, 0.432]	0.862	0.862	0.065	0.741

Table A.7: UEQ items of the OLS regression model fit on $zMCQ6$ with Spearman's ρ.

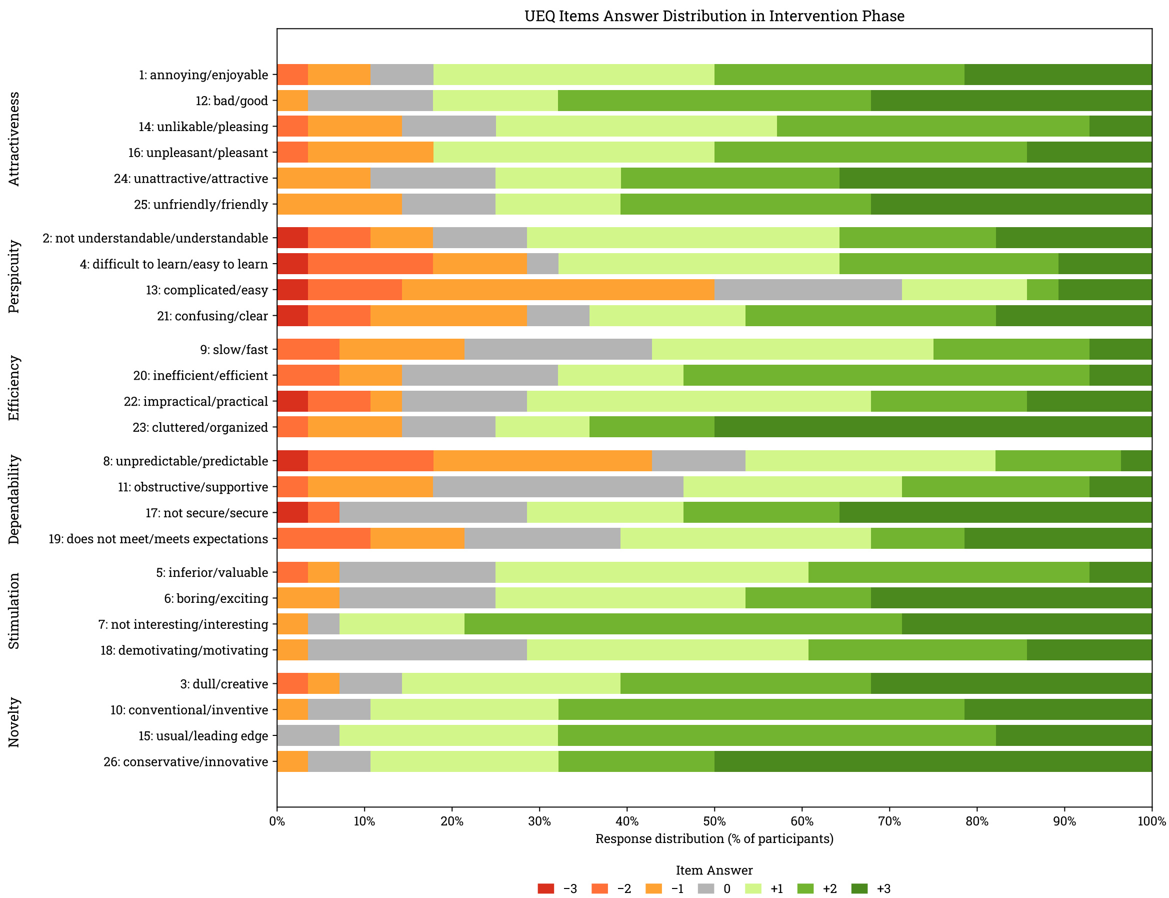


Figure A.2: The UEQ scale answer distribution per item for the Intervention phase.

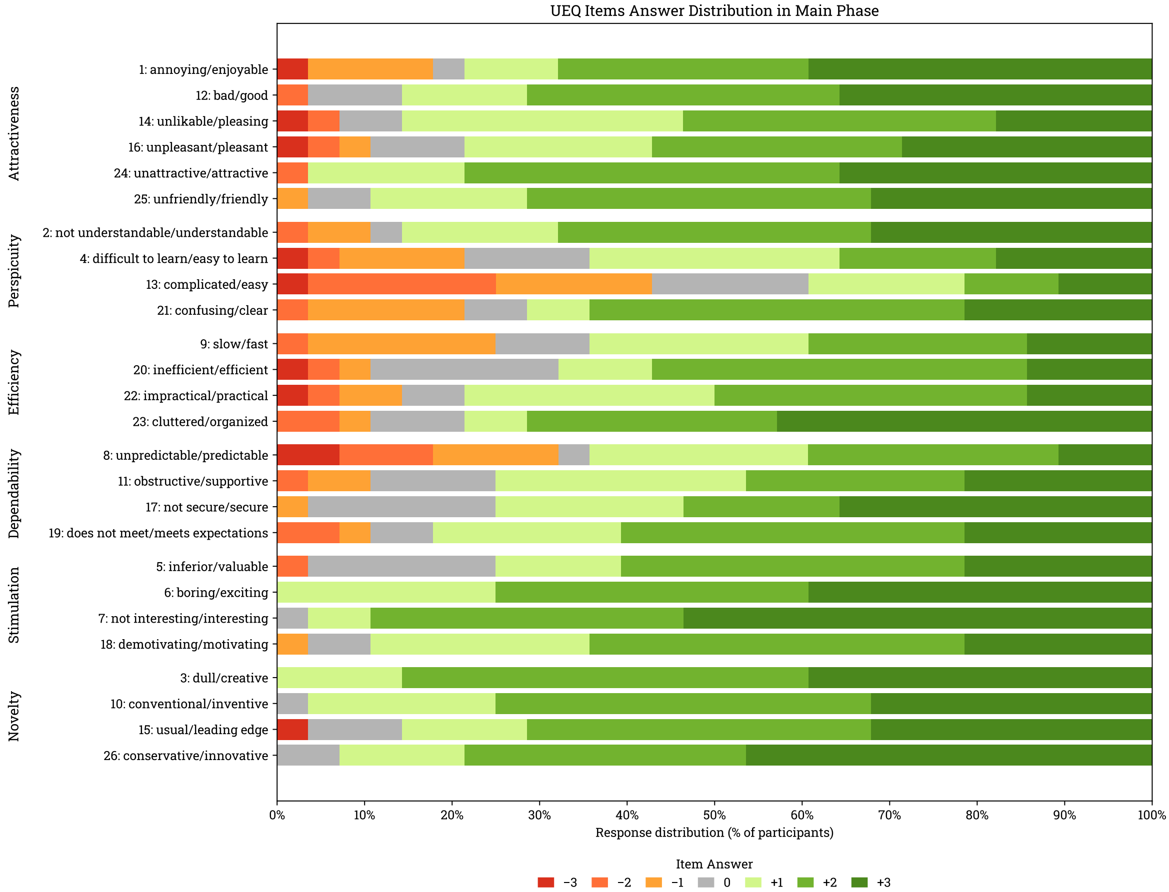
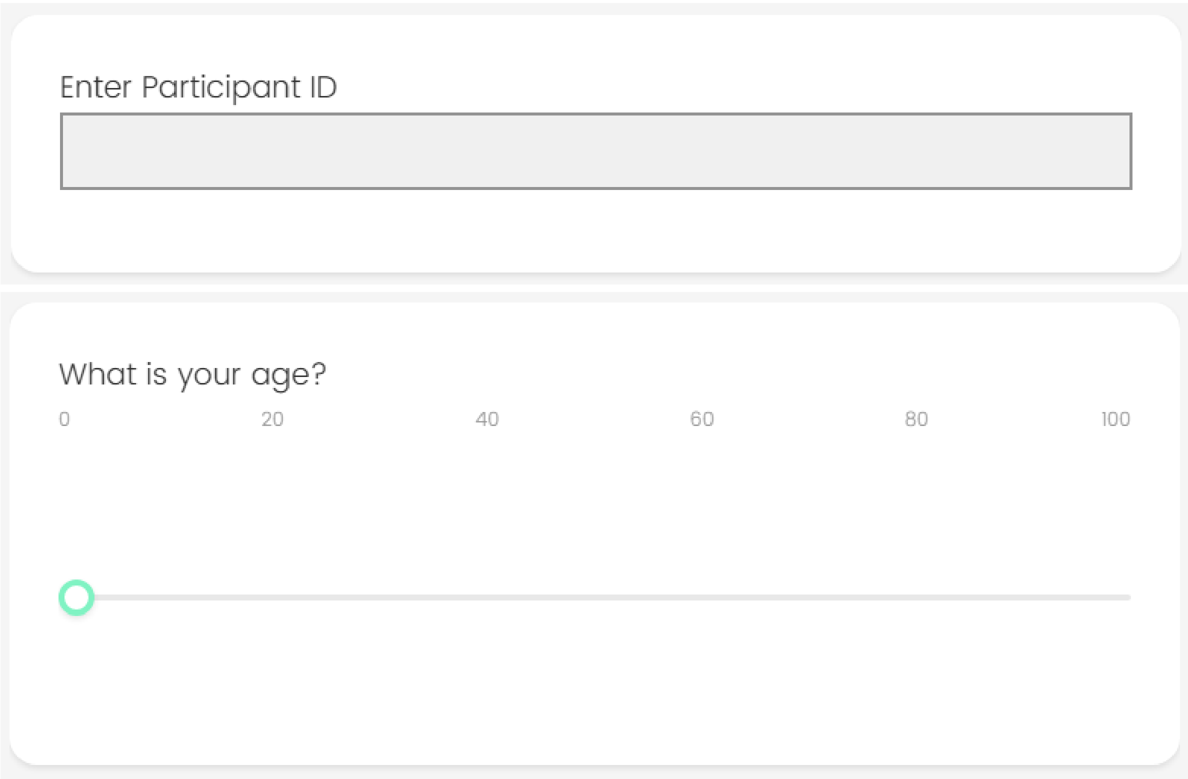


Figure A.3: The UEQ scale answer distribution per item for the Main phase.

B

Extra results

B.1. Participant Information



What is your gender?

- ◯ Male
- ◯ Female
- ◯ Non-binary / third gender
- ◯ Prefer not to say

What is your field of study?

- ◯ Generic programmes and qualifications
- ◯ Education
- ◯ Arts and humanities
- ◯ Social sciences, journalism and information
- ◯ Business, administration and law
- ◯ Natural sciences, mathematics and statistics
- ◯ Information and Communication Technologies
- ◯ Engineering, manufacturing and construction
- ◯ Agriculture, forestry, fisheries and veterinary
- ◯ Health and welfare
- ◯ Services

What is your educational attainment?

- ◯ High school
- ◯ Bachelor Degree
- ◯ Master Degree
- ◯ Doctoral Degree
- ◯ Other

How experienced are you with video games?

- ◯ Not experienced
- ◯ Slightly experienced
- ◯ Moderately experienced
- ◯ Very experienced
- ◯ Extremely experienced

How experienced are you with virtual reality (VR)?

- ◯ Not experienced
- ◯ Slightly experienced
- ◯ Moderately experienced
- ◯ Very experienced
- ◯ Extremely experienced

How familiar are you with hyperbolic geometry?

- ◯ Not familiar
- ◯ Slightly familiar
- ◯ Moderately familiar
- ◯ Very familiar
- ◯ Extremely familiar

In your everyday life (outside of VR), how often do you experience feelings of claustrophobia in situations like elevators, small rooms, tunnels, or crowded places?

- ○ Never
- ○ Rarely (a few times a year or less)
- ○ Occasionally (a few times a month)
- ○ Frequently (a few times a week)
- ○ Very frequently (almost daily or in most enclosed spaces)

B.2. UEQ (Intervention and Main)

The UEQ for the Main phase was excluded for redundancy

Please make your first evaluation now.

For the assessment of the intervention you just experienced, please fill out the following questionnaire. The questionnaire consists of pairs of contrasting attributes that may apply to the experience. The circles between the attributes represent gradations between the opposites. You can express your agreement with the attributes by ticking the circle that most closely reflects your impression.

Example:

Attractive ○ ● ○ ○ ○ ○ ○ Unattractive

This response would mean that you rate the application as more attractive than unattractive.

> Please decide spontaneously. Don't think too long about your decision to make sure that you convey your original impression.
>
> Sometimes you may not be completely sure about your agreement with a particular attribute or you may find that the attribute does not apply completely to the particular application. Nevertheless, please tick a circle in every line.

It is your personal opinion that counts. There is no wrong or right answer! Please answer ONLY for the experience in the intervention, NOT for any other parts of this study

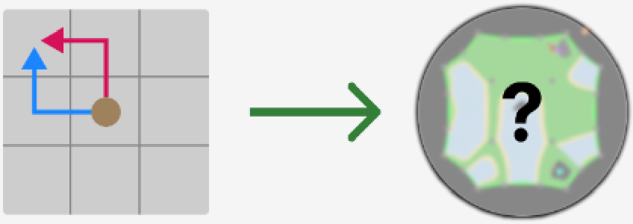
Annoying	Enjoyable
Not understandable	Understandable
Creative	Dull
Easy to learn	Difficult to learn
Valuable	Inferior

Boring	○	○	○	○	○	○	○	Exciting
Not interesting	○	○	○	○	○	○	○	Interesting
Unpredictable	○	○	○	○	○	○	○	Predictable
Fast	○	○	○	○	○	○	○	Slow
Inventive	○	○	○	○	○	○	○	Conventional
Obstructive	○	○	○	○	○	○	○	Supportive
Good	○	○	○	○	○	○	○	Bad
Complicated	○	○	○	○	○	○	○	Easy
Unlikable	○	○	○	○	○	○	○	Pleasing
Usual	○	○	○	○	○	○	○	Leading edge
Unpleasant	○	○	○	○	○	○	○	Pleasant
Secure	○	○	○	○	○	○	○	Not secure
Motivating	○	○	○	○	○	○	○	Demotivating
Meets expectations	○	○	○	○	○	○	○	Does not meet expectations
Inefficient	○	○	○	○	○	○	○	Efficient
Clear	○	○	○	○	○	○	○	Confusing

Impractical	○	○	○	○	○	○	○	Practical
Organized	○	○	○	○	○	○	○	Cluttered
Attractive	○	○	○	○	○	○	○	Unattractive
Friendly	○	○	○	○	○	○	○	Unfriendly
Conservative	○	○	○	○	○	○	○	Innovative

B.3. Post-study questions

If you move a tile left and then a tile up, will you end up in the same location (in the Hyperbolic world) if you move a tile up and then left?



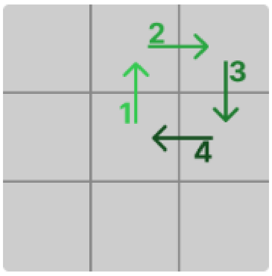
○ Yes, you end up in the same location

◉ No, you end up in a different location

How many tiles, exist around a tree in the hyperbolic world?

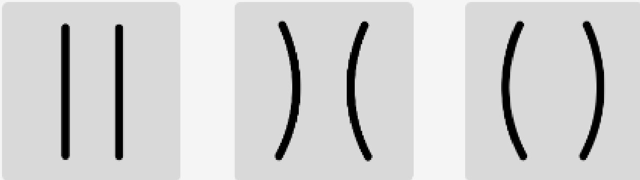
- ◯ 3
- ◯ 4
- ◉ 5
- ◯ 6

Moving (1)Up → (2)Right → (3)Down → (4)Left, as seen in the figure below, causes:



- ◯ You to end up one tile ahead if your starting position
- ◉ The (hyperbolic) world to rotate 90° clockwise
- ◯ A random teleport
- ◯ Nothing, you return to the same location and orientation

Imagine placing two points side by side and pointing them in the same direction. If both points move forward in that direction in the hyperbolic world, what will happen to the paths they follow?



(a) Parallel **(b) Diverge** **(c) Converge**

- (a) The paths move parallel to each other (remain the distance apart)
- (b) The paths diverge (move away from each other)
- (c) The paths converge (will cross eventually)

Roughly how much time do you think you spent interacting with the environment (trees, bushes, but not switches) in Holonomy?

- Chose not to
- Less than 30 seconds
- Between 30 seconds and 1 minute
- Between 1 minute and 2 minutes
- Between 2 minutes and 3 minutes
- More than 3 minutes

How engaging or memorable did you find the interactions with the environment?

- ◯ Not applicable
- ◯ Not engaging at all
- ◯ Slightly engaging
- ◯ Moderately engaging
- ◯ Very engaging
- ◯ Extremely engaging

Please rate how much you agree with the following statements based on your experience in the Holonomy environment.

	Strongly disagree	Slightly disagree	Neutral	Slightly agree	Strongly agree
I felt like the space around me was closing in.	○	○	○	○	○
I had the urge to leave the virtual environment.	○	○	○	○	○
I felt trapped or stuck in the Holonomy space.	○	○	○	○	○
The environment felt too tight or confining.	○	○	○	○	○
I felt comfortable with the amount of space around me.	○	○	○	○	○

C
Informed consent form

Informed Consent Form

Haptic Feedback in Non-Euclidean VR Spaces: A Study with SenseGloves in Holonomy

Thesis Author

Andreas Achilleos, MSc Computer Science student, WIS, TUDelft
(a.achilleos@student.tudelft.nl)

Research Supervisors

Dr. Martin Skrodzki, Tenured Assistant Professor, CGV, TUDelft (m.skrodzki@tudelft.nl)
Dr. Marcus Specht, Professor, WIS, TUDelft (m.m.specht@tudelft.nl)

Institution

Delft University of Technology (TU Delft)

You are being invited to participate in a study investigating the effects of haptic hardware (Sense Gloves Nova 1) on navigating a Non-Euclidean Virtual Reality environment. The study involves using a VR headset (Meta Quest 2) and a pair of haptic feedback gloves to experience an onboarding intervention, followed by a user experience related survey (UEQ), completing exploration tasks in the 'Holonomy' environment, and finally answering another UEQ along with some questions to related to your understanding of the experiment. The study takes place at TU Delft facilities.

Experiment Process

- After being familiarized and equipped with the hardware, you will complete a **short (seated) session** aimed to help with understanding navigation in the non-Euclidean environment of Holonomy.
- You will then **fill in a questionnaire** assessing your overall user experience with the system thus far.
- You will then **complete 2 levels in Holonomy with increasing difficulty** that require you to navigate to, and activate pre-generated switches in the environment. The position of these switches is indicated on your minimap that updates as you move through the environment. You will physically move in the confines of a 9-tile square grid; moving across tiles updates the minimap.
- Once all tasks are completed, you will **fill in a questionnaire** assessing your overall user experience with the system, and **answer questions** regarding the understanding of the experiment.
- The total duration of the study is approximately **60-70 minutes**.

Collected Data

For this study, various data regarding your experience and performance will be collected: *completion times, timestamped navigational data, interaction data, age, previous experience with VR/video games, claustrophobic tendencies, questionnaire responses.*

- Your responses/performance data will be anonymized and stored on secure TU Delft storage.
- Data will be stored for up to five years for research purposes in compliance with TU Delft's policies.
- Access to the study's data is restricted to affiliated authorized researchers.
- Personally Identifiable Information (PII) will **not** be published. A random identifier is assigned for solely administrative purposes and further analysis of the data.
- Only aggregated results will be shared in the final research publication.

Participation Risk

The study heavily involves virtual reality and haptic hardware. A number of individuals may be susceptible to VR sickness and could exhibit symptoms of eye fatigue, headache, nausea, excessive sweating, disorientation. The environment in Holonomy may also induce claustrophobia due to the compact representation in the confines of the grid you will be moving in.

- If at any point you experience discomfort, you may choose to withdraw your participation.
- You can schedule or take breaks at any point should any discomfort occur.
- You may ask for clarification on any processes/parts of the study that are unclear.
- The haptic hardware will provide force feedback and tactile sensations. The hardware can only restrict your movements to induce these sensations but will **not** move your hands in any unpredictable harmful way.

Voluntary Participation

- Participation in this study is entirely voluntary.
- You may choose to withdraw at any point with no explanation or consequences.
- You may request the deletion of your data upon withdrawal.

Please check all appropriate fields:

- ☐ I confirm that, I have read, understood, and had the opportunity to ask questions regarding the given information.
- ☐ I understand my participation is voluntary and I have the right to withdraw at any point with no consequences.
- ☐ I understand the role of my participation and the purpose of this study.
- ☐ I consent to the use of my anonymized data for research purposes.
- ☐ I understand that no personally identifiable information is linked to my participation data.
- ☐ I understand the possible risks associated with my participation in this study and the measures taken to mitigate them.
- ☐ I understand that my anonymized performance and questionnaire data will be tracked and analysed.
- ☐ I understand that my data can be withdrawn before the research publication.

____________________	____________________	____________________
Name of participant	**Signature**	**Date**

I, as a researcher, have accurately informed the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

Andreas Achilleos	____________________	____________________
Researcher name	**Signature**	**Date**