TITLE

Evaluation of spatial learning and wayfinding in a complex maze using immersive virtual reality. A registered report.

Eudave, L.[1*], Martínez M.[1], Valencia M.[2,3,4], Roth D.[5]

[1] Faculty of Education and Psychology, University of Navarra, Pamplona, Spain.

[2] Biomedical Engineering Program, Universidad de Navarra, CIMA, Pamplona, Spain

[3] IdiSNA, Navarra Institute for Health Research, 31080, Pamplona, Spain.

[4] Institute of Data Science and Artificial Intelligence, Universidad de Navarra, Pamplona, Spain.

[5] Faculty of Engineering, Friedrich-Alexander University, Erlangen, Germany.

*Corresponding Author: Dr Luis Eudave, University of Navarra, Pamplona, Spain, leudave@unav.es

CRediT Statement

Conceptualization: L.E; **Data curation:** L.E; **Formal Analysis:** L.E; M.M.; **Funding Acquisition:** L.E; **Investigation:** L.E; M.M; **Methodology:** L.E; M.M; M.V; D.R; **Project Administration:** L.E; M.M; **Resources:** L.E; M.M; **Software:** L.E; D.R; **Validation:** L.E; M.M; **Visualisation:** TBC. **Writing - original draft:** L.E; **Writing – review & editing:** L.E; M.M; M.V; D.R.

Abstract

Objectives: Mazes have traditionally been used as tools for evaluating spatial learning and navigational abilities in humans. They have been also utilized in sleep and dream research, as wayfinding is a common dream theme and participants undergoing experiments in the laboratory often dream about it. One such maze is the virtual maze task (VMT) created by Wamsley et al. (2010) to study the impact of sleep and dreaming in learning. Despite positive results found in several of those studies (dreaming of the VMT improves task performance), others failed to replicate these findings, possibly due to intrinsic methodological difficulties such as low task incorporation in dreams and the presence of cybersickness symptoms during task execution. It is possible that by using an adequately designed immersive virtual reality experience, which allows for a more naturalistic, stimulating and engaging simulation, these handicaps can be overcome. This Registered Report therefore aims to reproduce the original VMT version and compare it with an immersive virtual reality (iVR) adapted version using several wayfinding performance dependent measures. **Methods:** A sample of 58 participants were randomly allocated to one of the two conditions (Desktop vs. iVR) of the VMT task, where we measured performance and path variables. They then completed self-report measures of cybersickness symptoms, sense of presence during the task and a test for the assessment of spatial learning. **Results:** [TBD]. **Conclusions:** [TBD].

Key words: spatial learning; wayfinding; navigation; virtual reality; maze; presence; cybersickness.

Introduction

Navigation is an essential cognitive skill that humans use to orient themselves in space and move through the environment. The discovery of distinct specialized neurons has provided new insights into the neural mechanisms needed for successful navigation. Generally speaking, these mechanisms rely on a combination of sensory inputs, memory, and spatial reasoning. To navigate, humans can learn about the environment through visual cues such as landmarks and its layout, as well as auditory and tactile cues, allowing them to infer novel paths (place strategy). Alternatively, navigators can rely on their own position and orientation in relation to their surroundings to determine their route (response strategy), like moving towards a mountain (Ekstrom et al., 2018).

Traditionally, studies that explore navigational and wayfinding abilities in humans rely on the use of different kinds of mazes. Examples include T-mazes, plus sign mazes, star mazes or the computer based (virtual) version of the Morris water maze (VMWM) which has been used in up to a third of the human navigation studies (Thornberry et al., 2021). Wayfinding can also be assessed using virtual hedge mazes or city/urban mazes, where the main objective is to move from one point to another, usually making use of landmarks, from an egocentric or first person point of view. Despite some limitations associated with the use of mazes as a research tool, such as differences in experimental procedures, maze size, or measurements taken, they have proven valuable for gaining a better understanding of how navigation works in humans. Moreover, they allow for a more controlled and systematic environment, where it is possible to obtain valid and reliable data (Spiers et al., 2023). Nevertheless, most of these maze tasks still lack ecological validity as they are primarily presented through a computer screen.

One example of a frequently utilized maze in the literature is the Virtual Maze Task (VMT) created by Wamsley and colleagues to evaluate the effect of dreaming in a navigation task

(Murphy et al., 2018; Nguyen et al., 2013; Stamm et al., 2014; Wamsley et al., 2016; Wamsley, Tucker, Payne, & Stickgold, 2010; Wamsley, Tucker, Payne, Benavides, et al., 2010; Wamsley & Stickgold, 2019). Wayfinding and other exploratory behaviors are a common theme when dreaming (Yu, 2016a, 2016b), especially when participants take part in dream studies (Picard-Deland et al., 2021). The maze consists of a visually sparse (Wamsley, Tucker, Payne, & Stickgold, 2010) 20x20 grid, containing alleys, rooms and intersections with up to 5 different options. It includes two different types of landmarks as well as fog to force participants to rely on proximal visual information rather than visual cues in the distance (Nguyen et al., 2013). These design features were included to avoid memorization of a particular path or other response-based navigation strategies and instead rely on hippocampal place-response spatial strategies (Wamsley, Tucker, Payne, & Stickgold, 2010; Wamsley, Tucker, Payne, Benavides, et al., 2010) and the formation of cognitive maps (Nguyen et al., 2013). During the test phase of the VMT, participants are instructed to find their way to the exit as quickly as possible after a short training period. While originally devised to test the effect of task incorporation during sleep, it has also explored the role of experience (Wamsley, Tucker, Payne, & Stickgold, 2010), reward and sensory feedback (Stamm et al., 2014), and expectation (Wamsley et al., 2016) in VMT performance.

The main result from the first study of this series was that participants who slept and dreamt about the task performed better than those who slept but didn't dream about the task or those evaluated during the day (Wamsley, Tucker, Payne, Benavides, et al., 2010). This effect has been corroborated in later studies (Nguyen et al., 2013; Wamsley, Tucker, Payne, & Stickgold, 2010; Wamsley & Stickgold, 2019), while others have failed to replicate it (Murphy et al., 2018; Stamm et al., 2014; Wamsley et al., 2016). The authors have attributed this lack of consistency to the small number of participants who incorporated the VMT into

their dreams, the collection of dreams only during NREM sleep, and the use of modified versions of the original VMT (Wamsley & Stickgold, 2019).

One solution that could improve the incorporation rate into dreams is the use of immersive virtual reality (iVR) as a medium for displaying navigational tasks. Over the past decade, we have witnessed a renewed advent of iVR, where technological advancements and lower production costs have enabled its adoption by the general public and the research community. In this context, iVR has enabled the study of many cognitive processes in a more naturalistic and interactive manner, allowing for different ways of tracking behavioral responses, while simultaneously achieving highly controlled experimental procedures (Pan & Hamilton, 2018; Slater & Sanchez-Vives, 2016). In the case of spatial cognition research, iVR is able to combine the sensory flow generated by the virtual environment with the movements of the navigator, creating experiences that closely resemble those in the real world and beyond (Diersch & Wolbers, 2019). Consequently, the use of iVR to enhance ecological validity in navigation tasks has proven to be advantageous. Although a majority of studies indicate no differences in performance between Desktop and iVR versions of navigation tasks (Carbonell-Carrera et al., 2021; Clemenson et al., 2020; Hejtmanek et al., 2020; Zhao et al., 2020), iVR paradigms appear to result in greater real-world spatial knowledge transfer compared to less immersive alternatives (Hejtmanek et al., 2020). Moreover, specific populations may benefit from more naturalistic navigation tasks. When navigating in iVR, improvements in performance have been demonstrated in older adults who typically struggle with Desktop or 2D navigation (Hill et al., 2023; Ijaz et al., 2019; Wenk et al., 2022) or have restrictions in mobility (Diersch & Wolbers, 2019).

Within sleep and dream research, it has been demonstrated that enhanced or novel sensory stimulation using iVR during a given day increases the probability of dreaming about it overnight. For instance, training in lucid dreamers using iVR led to stronger increases in

6

dream lucidity compared to classic or no training (Gott et al., 2021). Similarly, iVR flying also resulted in more flying dreams when compared to the non-iVR control condition (Picard-Deland et al., 2020). This finding can be partially explained by the fact that high emotional intensity in waking-life experiences is a good predictor for dream incorporation (Malinowski & Horton, 2014; Schredl, 2006), and iVR experiences have been proven to elicit strong emotional arousal (see Somarathna et al., 2022 for a review), even when compared to 2D experiences, both subjectively and physiologically (Ding et al., 2018; Flavián et al., 2021; Tian et al., 2021, 2022; Xie et al., 2023). Given that current iVR devices provide a more immersive, salient and emotionally engaging environment, it is plausible that by solving an iVR version of the VMT, incorporation during dreaming becomes more likely.

Another important issue present in the studies conducted by Wamsley and colleagues was the exclusion of participants based on their previous exposure to video games. This approach may lead to biased results that are not fully representative of the sampled population. One argument used to justify this decision was the poor performance and failure to demonstrate sleep-dependent improvement, as well as the higher incidence of cybersickness symptoms in these "novice" participants (Nguyen et al., 2013). Given that participants who performed poorly during training showed greater overnight improvements (Wamsley, Tucker, Payne, Benavides, et al., 2010; Wamsley & Stickgold, 2019) it is pertinent to reconsider retaining "novice" participants for evaluation, as they may be the subgroup that benefits the most. If this is indeed the case, then it becomes important to reduce or alleviate the number and intensity of cybersickness symptoms.

When designing a three-dimensional experience like the VMT, it is essential to incorporate measures that reduce the risk of eliciting cybersickness, especially for iVR. Successfully implemented examples of such measures include peripheral blurring or field of view occlusion (Groth et al., 2021; refer to Method section for more details). Moreover, it has been

observed that a negative correlation exists between the incidence of cybersickness symptoms and the sense of presence (the feeling or illusion of "being there") within the virtual experience (Martirosov et al., 2022; Thorp et al., 2022; Weech et al., 2019). Since iVR experiences tend to evoke a higher sense of presence, it is crucial for researchers to consider certain environmental and user design parameters that can help enhance this feeling, such as visual and audio fidelity, visual and sensorimotor feedback, and so on. Overall, taking these interventions into account has been proven beneficial in reducing cybersickness symptoms and improving the overall experience, which, in turn, can help retain at least a number of cybersickness-prone participants.

Study Overview & Hypotheses

Here, we propose creating an adaptation of the VMT that is suitable for use in iVR. We aim to compare it in terms of performance, cybersickness symptoms and presence, to the traditional desktop version of the VMT. This will ensure that both tasks can effectively measure the spatial learning and navigational variables of interest while also allowing the creation of a more ecological and immersive task when using iVR.

We expect to obtain equivalent results in terms of performance between the desktop and iVR versions of the VMT. Previous studies evaluating navigation and spatial learning between iVR and non-iVR systems have generally found no significant differences between the two groups (Aoki et al., 2008; Barrett et al., 2022; Carbonell-Carrera et al., 2021; Carbonell-Carrera & Saorin, 2017; Marraffino et al., 2022; Weidner et al., 2017; Zhao et al., 2020), while some have reported mixed results (Feng et al., 2022; Murcia-López & Steed, 2016; Sousa Santos et al., 2009; Srivastava et al., 2019). Additionally, we intend to openly share our adaptation of the VMT, including its project code and materials to promote

transparency and replicability, as has been done with other maze types (Commins et al., 2020).

Concretely, we will attempt to answer the following research questions:

RQ1: Is there a difference in spatial learning performance between the Desktop and iVR versions of the VMT?

RQ2: Are there differences between Desktop and iVR versions of the VMT in terms of presence and cybersickness symptoms?

Based on the review of previous research studies that use and compare Desktop and iVR tasks, we have the following hypotheses for our research questions:

H1: Spatial learning performance will be equivalent in both Desktop and iVR versions of the VMT.

H2: In the iVR version of the VMT, cybersickness symptoms incidence will be equivalent or lower and presence will be higher when compared to the Desktop version.

Method

Participants

This study was approved by the University of Navarra Ethics Committee (project number 2022.090). This study employs a two-group (Desktop vs. iVR) between-subjects design. To be eligible for the study, participants must confirm that they are between 18 and 30 years old and do not have a history of neurological or psychiatric diagnoses. During the sign-up process, participants will be asked about their prior exposure to videogames and/or 3D experiences. Participants will be recruited through internal communication channels of the authors' institution and social media platforms such as Twitter or Instagram. No compensation will be provided to participants for their involvement in this study.

In some of the studies that used the VMT, participants were excluded if they had no experience with 3D-style videogames (less than once per year), since it was discovered that these subjects exhibited lower overall performance in the VMT (Wamsley, Tucker, Payne, & Stickgold, 2010), along with a higher incidence of cybersickness symptoms (Nguyen et al., 2013). To incorporate these participants in our study, we will invite those who indicate having less than once per year of videogame/3D experience to the laboratory for a 5-minute iVR scenario exposition, specifically the Adaptation phase from the VMT (details provided below). If no cybersickness symptoms are reported during this period, they will be enrolled in the complete procedure.

Our calculated sample size is based on the effect sizes reported in the studies by Wamsley et al. (2010; 2019). As in these studies, our designated main effect of interest is completion time improvement, where Wamsley et al. (2010) observed a significant effect size of $d = 2.2$ and Wamsley et al. (2019) found an effect size of $d = 1.1$. From these studies, we selected the one with the lowest effect size, as recommended by Lakens et al. (Lakens et al., 2022). Following

a *small telescopes* approach (Simonsohn, 2015), we set our smallest effect size of interest (SESOI) as the effect size that the aforementioned study would have had 33% power to detect which is $d = 0.77$. Using this SESOI, we conducted a series of sensitivity power analyses based on the two one-sided tests procedure for equivalence testing (TOST, Lakens, 2017). This resulted in a sample size of 58 participants (29 subjects per group), which is required to achieve 80% statistical power with alpha set at .05.

It is important to note that the main effect of interest, which was used to estimate our sample size, was extracted from a two-group comparison that will not be evaluated in this study. It involves comparing completion time improvements between sleeping and awake participants (who either dreamt or not about the VMT), which is not the focus of the current study. Nevertheless, we believe that this method can help us validate our VMT adaptation for a future replication study of the aforementioned studies.

Participants will be randomly allocated to one of the two conditions: the desktop version of the VMT (D-VMT) or the immersive virtual reality version of the VMT (iVR-VMT). Specifically, we will use a stratified permutation block randomization with two strata for sex to ensure an equal number of males and females in each group..

Hardware

For the iVR-VMT group, we will use a Meta Quest 2 head mounted display (HMD) with a horizontal field of view (FOV) of 89±4º and a 90 Hz refresh rate, along with two controllers will be used for navigation and motion tracking within the virtual environment. The iVR area has a size of 5m2, which is adequate for our seated iVR experience. The HMD will be connected to a laptop with an Intel Core i7 7700HQ 2.80 GHz processor, 16 GB RAM, a 4095 MB NVIDIA GeForce GTX 1070 graphics card, a 931 GB TOSHIBA MQ01ABD100

(SATA) hard disk, and Realtek High Definition Audio. This laptop will wirelessly stream the PC-native program running the VMT using a router (TP-Link AX10) via the VirtualDesktop application available for both the PC and Meta Quest 2. Participants in the D-VMT condition will use the same laptop connected to an external keyboard and a 27-inch monitor with a display resolution of 1024x768, seated from a distance of 60cm (resulting in a field of view ~50º). This resolution was chosen to preserve the aspect ratio used in previous studies.

Materials

Virtual Maze Task

The Virtual Maze Task (VMT) that will be used in this experiment is a reproduction of the maze utilized in prior research (Murphy et al., 2018; Stamm et al., 2014; Wamsley et al., 2016). This encompasses its layout, the 20x20 grid design, landmark locations, as well as entry and exit points (Figure 1 and Supplementary Figure 1). In general, it consists of alleys, squares and dead ends filled with two types of landmarks (palm trees and floor lights). Fog mechanics were implemented so that visibility diminishes linearly from the participant's point of view. It was designed and developed using the Unity game engine (v2021.1.14f1). All project files will be available as a Github repository (https://github.com/negatoscope/VRMaze) along with its latest standalone build on OSF (https://osf.io/2g6b8).

Procedurally, our version of the VMT consists of three phases. First, in the Adaptation phase, participants appear in a smaller (unrelated to the task) version of the maze, consisting of two rooms connected by an alley. Participants are instructed to learn the controls (arrow keys for D-VMT and direction of the head + right index trigger to move forward for iVR-VMT) and to get familiarized themselves with the environment and the landmarks. We will also take this opportunity to screen for potential cybersickness symptoms in susceptible participants, which

could impede task completion. This phase has a duration of 5 minutes (participants could end it earlier by reaching the exit door). Next, a 5 minute Training phase ensues, where participants spawn next to the exit door and are instructed to explore the maze as much as possible while trying to memorize the way to the exit, as they will to be tested afterward. After 5 minutes, the Trial phase begins, comprising 3 trials in which participants attempt to reach the exit as quickly as possible. Due to the maze's layout, there are several paths that lead to the exit with "optimal paths" requiring 79 grids or 191 in-game units (Figure 1). These "optimal paths" bifurcate at grid r16, creating two equidistant alternatives to the exit. In each trial, the participants will spawn at one of the 3 spawn locations (pseudorandomized at the beginning of the experiment, counterbalanced for all participants), facing a wall. These locations are equidistant from the exit and differ by no more than 1 second in completion time. Trials conclude when participants reach the exit door or after 10 minutes.

Controls in the D-VMT involve the use of the arrow keys (up arrow to move forward, left-right to rotate). In the iVR-VMT version, participants move forward by rotating the chair (or looking) in the desired direction while wearing the HMD and pressing the right control index trigger button to initiate movement (continuous locomotion). We chose this mode of iVR locomotion for two reasons: 1) it is more intuitive and easier to learn when compared other types (teleportation, snap turning, etc.) which could prevent our results from being influenced by a lack of skill in learning how to move in iVR; and 2) to enable us to compare performance parameters between both D-VMT and iVR-VMT versions (completion time, distance, speed, etc.), as participants in both will be able to move at the same speed. However, continuous locomotion is prone to induce more cybersickness symptoms (Saredakis et al., 2020) due to the apparent sensory mismatch between virtual displacement (perceived self-motion) while remaining stationary in the physical world.

To minimize the chances of participants experiencing symptoms of cybersickness, several adaptations were implemented in the VMT. The maximum forward movement was limited to 3 in-game units per second (roughly equivalent to 3 meters per second), following the recommendations in the Oculus VR Best Practices Guide (Yao et al., 2014). To mitigate cybersickness symptoms during rotations and translational movement, peripheral blurring and field of view (FOV) reduction were implemented in both the D-VMT and iVR-VMT, as studies have shown that these measures can be beneficial in scenarios involving continuous locomotion (Groth et al., 2021). The extent of peripheral blur and FOV occlusion were adjusted according to the intensity and duration of movement, with greater blur and occlusion applied during rapid rotations and/or forward movement.

Figure 1. Virtual Maze Task. Top left: layout (top-view) of the rendered 3D maze (participants were not shown this image). Top right: 20x20 grid layout of the VMT (L1: palm tree; L2: floor light; S1-3: starting points; E: exit). The optimal paths are represented as colored grids, and include paths starting from S1 (red), S2 (blue) and S3 (yellow; green when it joins S2 path) which then combine in a common path (purple). Solid grid colors represent one of the possible optimal paths, and lightly-colored grids represent possible alternative grids within an optimal path. Note the bifurcation in grid r16 that creates two equidistant alternatives to reach the exit. Bottom: in-game egocentric (first person) view of the Adaptation phase of the VMT showing the two types of landmarks (palm tree and floor light) and the exit door.

VMT variables

The primary dependent measure for the VMT (as established in previous studies) is Completion Time, defined as the number of seconds required to reach the maze's exit in each trial. Other dependent measures extracted from each trial include: Distance Traveled, defined as the number of grid boundaries traversed; Backtracking, defined as the proportion of retraced grids (1 − number of unique grid positions/Distance Traveled); and Moving Speed (Distance Traveled/Completion Time). Our main effect of interest, improvement, is calculated both (1) as the difference in the raw performance at retest versus pretest (mean performance on three retest trials − mean performance on three pretest trials), and (2) as the percentage of improvement relative to baseline performance (raw performance change/mean performance on pretest trials) for all dependant measures.

In a secondary analysis, we will assess the effect of gender and spatial learning skill on our participants' performance on the VMT, as both measures have shown to have an impact in maze and spatial learning (Coutrot et al., 2018; Mueller et al., 2008; Piber et al., 2018; Woolley et al., 2010).

Simulator Sickness Questionnaire (SSQ)

The adapted Spanish version of the SSQ (Campo-Prieto et al., 2021) will be used to assess the presence and intensity of cybersickness symptoms. The SSQ represents a widely used tool to measure cybersickness and consists of 16 items with a four-point Likert scale response (0-none, 1-slight, 2-moderate and 3-severe). These items are categorized into three dimensions or factors: oculomotor, disorientation and nausea. Four scores will be calculated, including one for each factor (maximum score: 100; the formula for its calculation is detailed in Kennedy et al., 1993) as well as a total score (maximum score: 300). Participants will be instructed to complete the SSQ before and after each VMT session.

Sense of Presence Inventory (ITC-SOPI)

We will utilize a Spanish translated version of the ITC-SOPI self-report questionnaire developed by Lessiter et al. (2001), developed to assess the subjective feeling of presence during and after exposure to different types of media. This questionnaire comprises 44 items that are rated using a five-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). Items are divided into four subscales: sense of physical space (PS), engagement (E), ecological validity (EV) and negative effects (NE). The scores for each subscale are calculated by summing each item's response.

Perspective Taking Test (PTT)

To assess our participants' spatial learning skill, we will employ the Perspective Taking Test. This test is the computarized version of the paper-based Spatial Orientation Test (SOT, Hegarty, 2004), which is a reliable and widely employed tool for evaluating the spatial ability of perspective taking (for a preview go to: https://negatoscope.shinyapps.io/PTT_SOT_spanish/; materials are available in https://github.com/negatoscope/PTT_spanish/). In essence, each trial consists of an array of objects, where the participant has to imagine being located at one of the objects, facing a second object. The objective is to estimate the direction of a third object by clicking on the circumference where a line appears indicating the chosen location (Friedman et al., 2020). Accuracy of responses is measured by the angular error (target angle - subject's selected angle).

Procedure

Participants will be asked to avoid having a heavy breakfast and drinking coffee (or any other stimulants) on the morning of the experiment. They will be requested to report to the laboratory at approximately 8:00 hours for the first session (S1). After signing the written

informed consent, they will respond to a questionnaire requesting information on demographics, computer, videogames and iVR use, as well as the SSQ. Participants will then be randomly assigned to either the D-VMT or iVR-VMT groups and will be given instructions on the controls and the Adaptation, Training and pretest Trial phases. If participants are assigned to the D-VMT they will sit on a comfortable desk chair in front of a desk with a keyboard and monitor located at 60cm from the participant. If they are assigned to the iVR-VMT, they will sit on the desk chair placed in the center of the iVR play area. After completing the VMT participants will respond to the SSQ and will be asked return for the second session (S2) at 13:00 hours. During S2, participants will answer the SSQ, retest the Trial phase of the VMT, and complete the SSQ once again, along with the ITC-SOPI and the PTT. Both sessions last approximately 45 minutes, for a total evaluation time of 90 minutes. The entire procedure is represented in Figure 2.
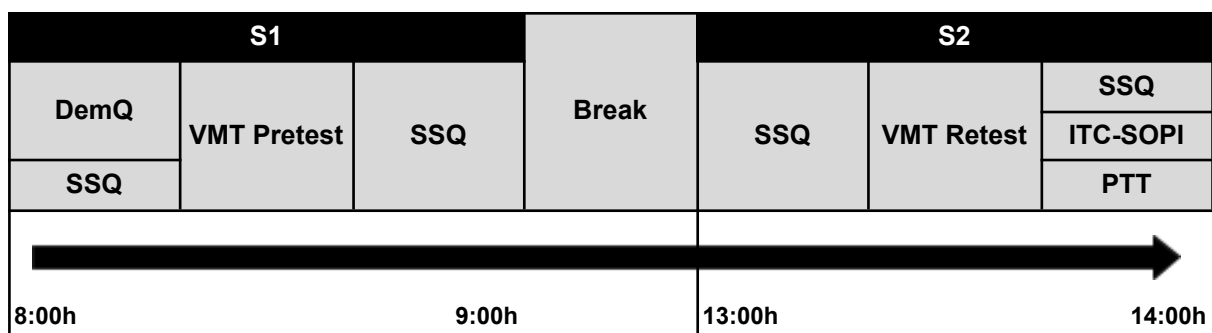

Figure 2. Experimental procedure. S1: Session 1; S2: Session 2; DemQ: demographics questionnaire; SSQ: Simulator Sickness Questionnaire; VMT: Virtual Maze Task; ITC-SOPI: ITC-Sense of Presence Inventory; PTT: Perspective-Taking Test.

Analytic Strategy

Independent samples equivalence tests (Lakens, 2017) will be conducted for each of our RQs. To compare our results to those obtained in the latest replication study by Wamsley et al. (2019), equivalence tests will be conducted on the measures of improvement (raw and

percentage) in Completion Time, Distance Travelled, Moving Speed and Backtracking. Equivalence tests use the two one-sided tests procedure to statistically reject the presence of effects large enough to be considered worthwhile. We will use the upper and lower equivalence bounds of $-\Delta_L = -.77$ and $\Delta_U = .77$ based on the previously calculated effect size that our design was sufficiently powered to detect. We set an alpha value at $\alpha = .05$ to denote statistical significance for p-values equal or lower to that threshold. Equivalence will therefore be asserted if, given $\alpha = .05$, the 99% confidence interval of the mean difference lies within this equivalence region, and rejected if the 99% CI lies outside of this region. Independent sample Welch's two-sided t-tests will be conducted for between-group comparisons for the subscale and total scores of the SSQ and ITC-SOPI questionnaires, as well as accuracy in the PTT. We set an alpha value at $\alpha = .05$ to denote statistical significance for p-values equal or lower to that threshold. To assess the effect of gender and spatial learning skill, linear regression will be used. Performance VMT variables will be considered the dependent variable, whereas gender and spatial learning skill (PPT accuracy score) will serve as independent variables. All statistical analyses will be carried out using R and RStudio. A summary for this analytic strategy can be found in Table 1.

Data and Code Availability

All materials, code and raw data will be made publicly available on the Open Science Framework: https://osf.io/j8qfv/. This study meets the Level 6 of the PCI RR bias control (https://rr.peercommunityin.org/help/guide_for_authors).
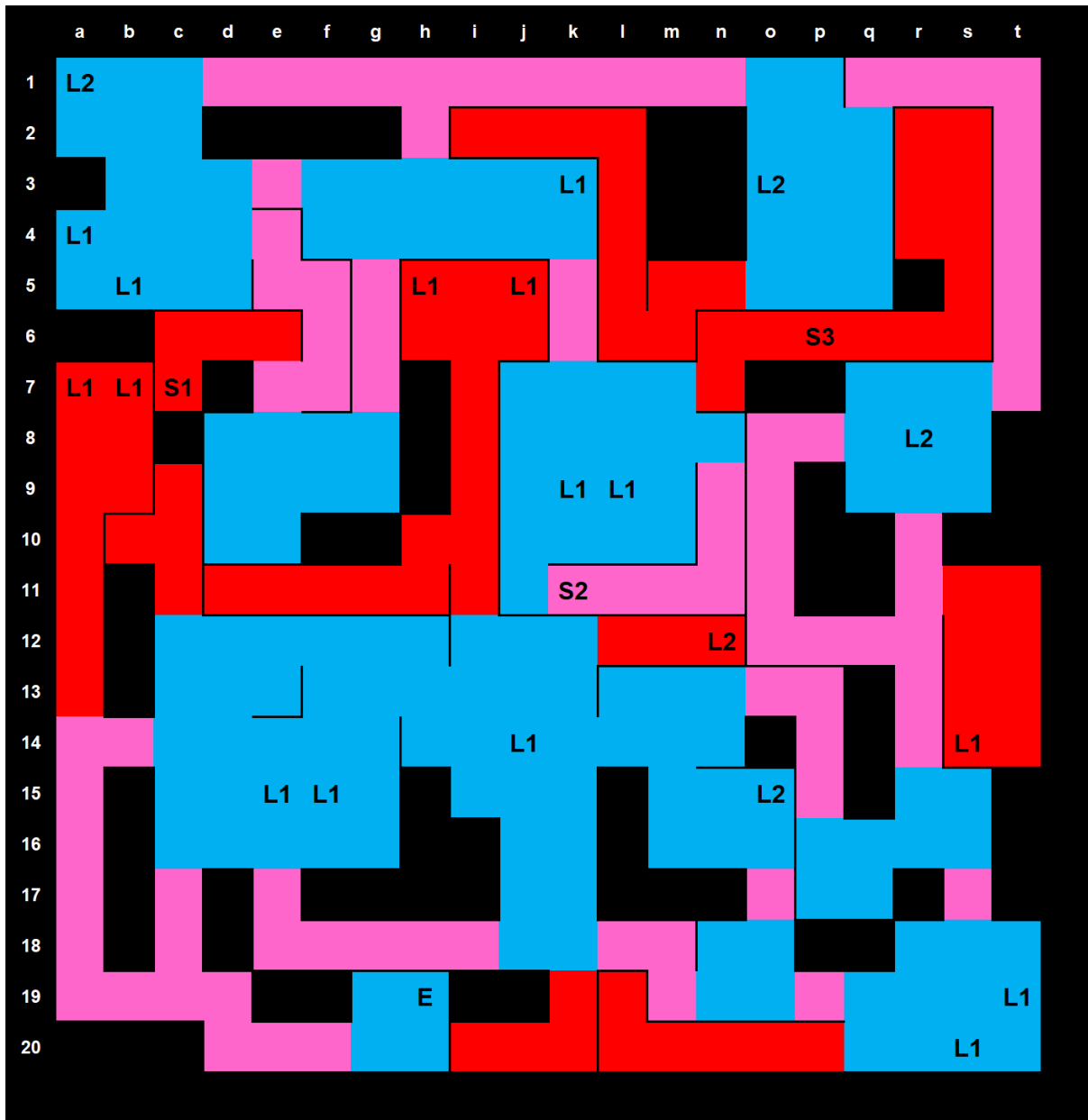
Results

TBD

Discussion

TBD

Table 1. Design summary.

Question	Hypothesis	Sampling plan	Analysis Plan	Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis	Interpretation given different outcomes	Theory that could be shown wrong by the outcomes
RQ1: Is there a difference in spatial learning between the Desktop and iVR versions of the VMT task?	Performance in the Desktop and iVR versions of the VMT task will be equivalent.	A sample size of 58 participants will be recruited, allowing for >80% statistical power and an alpha = 0.05 to conduct equivalence tests on the upper and lower equivalence bounds of $-\Delta L$ = -.71 and ΔU = .71.	Independent samples equivalence tests will be conducted on the between-participants 2 group comparison (Desktop vs. iVR) with the upper and lower equivalence bounds of $-\Delta L$ = -.71 and ΔU = .71.	If the 99% CI lies outside of the equivalence region ($-\Delta L$ = -.71 and ΔU = .71), we will assert a meaningful effect. If the 99% CIs lie within the equivalence region, we will assert that we did not detect a meaningful effect (given the effect size that our sample is powered to detect), and therefore are equivalent.	If we find evidence of a meaningful effect, then this will suggest that either condition (Desktop or iVR) is better at detecting spatial learning. If this effect is equivalent, then this will suggest that both conditions are similar in detecting spatial learning.	If we find evidence of a meaningful effect, this will contradict what some other studies have found when comparing desktop and iVR versions of other tasks, where no significant differences were found between versions (Barrett et al. 2022, Zhao et al. 2010)

| RQ2: Are there differences between Desktop and iVR versions of the VMT task in terms of presence and cybersickness symptoms? | Directional: We expect the iVR version of the VMT task to show better presence scores and equivalent or lower cybersickness symptoms. | As above. | Independent samples Welch's two-sided t-tests will be conducted on the between-participants 2 group comparison (Desktop vs. iVR). This analysis will be conducted on three discrete subscales of the Simulator Sickness Questionnaire (SSQ) and on the four discrete subscales of the ITC-SOPI Questionnaire. | If the resulting p-value is less than 0.05, we will assert a meaningful between-groups effect. | If we find evidence of a meaningful effect, then this will provide support for one of both versions having better usability (higher), presence (higher) and cybersickness symptoms (lower). | There is no sufficient evidence from studies comparing usability, presence or cybersickness between versions (Desktop vs. iVR) of the same task. However, due to the more naturalistic properties of the iVR version, we expect higher scores in usability and presence, and lower cybersickness symptoms. |
| | | | | | | |

Supplementary Figure 1. Virtual Maze Task layout. It consists of several open squares (blue), alleys (pink), dead ends (red), landmarks and an exit door. L1: palm tree; L2: floor light; S1-3: starting points; E: exit.

File S1: Simulator Sickness Questionnaire (Spanish adaptation by Campo-Prieto et al. 2022)

Señala la presencia/intensidad de los siguientes síntomas del 1 al 4, siendo:

- 0. Ausencia
- 1. Leve
- 2. Moderado
- 3. Grave

Síntomas:

- Malestar general
- Cansancio
- Dolor de cabeza
- Vista cansada
- Dificultad para enfocar
- Aumento de salivación
- Sudoración
- Náusea
- Dificultad para concentrarse
- Pesadez de cabeza
- Visión borrosa
- Mareos con ojos abiertos
- Mareos con ojos cerrados
- Vértigo
- Estómago revuelto
- Eructos

File S2: ITC-SOPI (Spanish translation)

Evalúa las siguientes expresiones del 1 al 5, siendo:

1. Totalmente en desacuerdo.
2. En desacuerdo.
3. Ni de acuerdo ni en desacuerdo.
4. De acuerdo.
5. Totalmente de acuerdo

Pregunta 1. TRAS mi experiencia en el entorno virtual...

- Me sentí triste porque la experiencia terminase.
- Me sentí desorientado.
- Tuve la sensación de que había vuelto de un viaje.
- Me hubiera gustado que la experiencia hubiera continuado.
- Recuerdo vívidamente algunas partes de la experiencia.
- Recomendaría esta experiencia a mis amigos.

Pregunta 2. DURANTE mi experiencia en el entorno virtual...

- Me sentí involucrado.
- Me sentí envuelto por el entorno.
- Perdí la noción del tiempo.
- Sentí que podía interactuar con el entorno.
- El entorno parecía natural.
- Sentí que el contenido estaba "vivo".
- Sentí que los elementos u objetos casi podían tocarme.
- Me divertí.
- Sentí que estaba visitando los lugares del entorno.
- Me sentí cansado.
- El contenido me pareció creíble.
- Sentí que no estaba sólo viendo cosas.
- Tuve la sensación de que me movía en respuesta a partes del entorno.
- Me sentí mareado.
- Sentí que el entorno era parte de la vida real.
- Mi experiencia fue intensa.
- Presté más atención al entorno que a mis propios pensamientos (preocupaciones personales, fantasías, ...)
- Tuve la sensación de estar dentro de las escenas del entorno.
- Sentí que podía mover objetos (en el entorno virtual).
- Las escenas representadas podrían ocurrir realmente en el mundo real.
- Sentí fatiga ocular.
- Casi podía oler los diferentes elementos del entorno.

- Tuve la sensación de que los elementos del entorno eran conscientes de que yo estaba allí.
- Tuve una fuerte sensación de que los sonidos provenían de diferentes direcciones del entorno.
- Me sentí rodeado por el entorno.
- Sentí náuseas.
- Tuve una fuerte sensación de que los elementos eran sólidos.
- Sentí que podía alcanzar y tocar cosas (en el entorno).
- Sentí que la temperatura cambiaba junto a las escenas del entorno.
- Respondí de manera emocional.
- Sentí que todos mis sentidos eran estimulados al mismo tiempo.
- El contenido me atrajo.
- Me sentí capaz de cambiar el curso de los eventos en el entorno.
- Sentí que estaba en el mismo espacio que los elementos del entorno.
- Tuve la sensación de que partes del entorno me respondían.
- Sentí que realmente podía mover cosas en el entorno.
- Sentí dolor de cabeza.
- Sentí que estaba participando dentro del entorno.

References

Aoki, H., Oman, C. M., Buckland, D. A., & Natapoff, A. (2008). Desktop-VR system for preflight 3D navigation training. *Acta Astronautica*, *63*(7–10), 841–847. https://doi.org/10.1016/j.actaastro.2007.11.001

Barrett, R. C. A., Poe, R., O'Camb, J. W., Woodruff, C., Harrison, S. M., Dolguikh, K., Chuong, C., Klassen, A. D., Zhang, R., Joseph, R. B., & Blair, M. R. (2022). Comparing virtual reality, desktop-based 3D, and 2D versions of a category learning experiment. *PLOS ONE*, *17*(10), e0275119. https://doi.org/10.1371/journal.pone.0275119

Campo-Prieto, P., Rodríguez-Fuentes, G., & Cancela Carral, J. M. (2021). Traducción y adaptación transcultural al español del Simulator Sickness Questionnaire (Translation and cross-cultural adaptation to Spanish of the Simulator Sickness Questionnaire). *Retos*, *43*, 503–509. https://doi.org/10.47197/retos.v43i0.87605

Carbonell-Carrera, C., Saorin, J., & Jaeger, A. (2021). Navigation Tasks in Desktop VR Environments to Improve the Spatial Orientation Skill of Building Engineers. *Buildings*, *11*(10), 492. https://doi.org/10.3390/buildings11100492

Carbonell-Carrera, C., & Saorin, J. L. (2017). Virtual Learning Environments to Enhance Spatial Orientation. *EURASIA Journal of Mathematics, Science and Technology Education*, *14*(3). https://doi.org/10.12973/ejmste/79171

Clemenson, G. D., Wang, L., Mao, Z., Stark, S. M., & Stark, C. E. L. (2020). Exploring the Spatial Relationships Between Real and Virtual Experiences: What Transfers and What Doesn't. *Frontiers in Virtual Reality*, *1*, 572122. https://doi.org/10.3389/frvir.2020.572122

Commins, S., Duffin, J., Chaves, K., Leahy, D., Corcoran, K., Caffrey, M., Keenan, L., Finan, D., & Thornberry, C. (2020). NavWell: A simplified virtual-reality platform for spatial navigation and memory experiments. *Behavior Research Methods*, *52*(3), 1189–1207. https://doi.org/10.3758/s13428-019-01310-5

Coutrot, A., Silva, R., Manley, E., De Cothi, W., Sami, S., Bohbot, V. D., Wiener, J. M., Hölscher, C., Dalton, R. C., Hornberger, M., & Spiers, H. J. (2018). Global Determinants of Navigation Ability. *Current Biology*, *28*(17), 2861-2866.e4. https://doi.org/10.1016/j.cub.2018.06.009

Diersch, N., & Wolbers, T. (2019). The potential of virtual reality for spatial navigation research across the adult lifespan. *Journal of Experimental Biology*, *222*(Suppl_1), jeb187252. https://doi.org/10.1242/jeb.187252

Ding, N., Zhou, W., & Fung, A. Y. H. (2018). Emotional effect of cinematic VR compared with traditional 2D film. *Telematics and Informatics*, *35*(6), 1572–1579. https://doi.org/10.1016/j.tele.2018.04.003

Ekstrom, A. D., Spiers, H. J., Bohbot, V. D., & Rosenbaum, R. S. (2018). *Human spatial navigation*. Princeton University Press.

Feng, Y., Duives, D. C., & Hoogendoorn, S. P. (2022). Wayfinding behaviour in a multi-level building: A comparative study of HMD VR and Desktop VR. *Advanced Engineering Informatics*, *51*, 101475. https://doi.org/10.1016/j.aei.2021.101475

Flavián, C., Ibáñez-Sánchez, S., & Orús, C. (2021). Impacts of technological embodiment through virtual reality on potential guests' emotions and engagement. *Journal of Hospitality Marketing & Management*, *30*(1), 1–20. https://doi.org/10.1080/19368623.2020.1770146

Friedman, A., Kohler, B., Gunalp, P., Boone, A. P., & Hegarty, M. (2020). A computerized spatial orientation test. *Behavior Research Methods*, *52*(2), 799–812. https://doi.org/10.3758/s13428-019-01277-3

Gott, J., Bovy, L., Peters, E., Tzioridou, S., Meo, S., Demirel, Ç., Esfahani, M. J., Oliveira, P. R., Houweling, T., Orticoni, A., Rademaker, A., Booltink, D., Varatheeswaran, R., van Hooijdonk, C., Chaabou, M., Mangiaruga, A., van den Berge, E., Weber, F. D., Ritter, S., & Dresler, M. (2021). Virtual reality training of lucid dreaming. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1817), 20190697. https://doi.org/10.1098/rstb.2019.0697

Groth, C., Tauscher, J.-P., Heesen, N., Castillo, S., & Magnor, M. (2021). Visual Techniques to Reduce Cybersickness in Virtual Reality. *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 486–487. https://doi.org/10.1109/VRW52623.2021.00125

Hegarty, M. (2004). A dissociation between mental rotation and perspective-taking spatial abilities.

Intelligence, *32*(2), 175–191. https://doi.org/10.1016/j.intell.2003.12.001

Hejtmanek, L., Starrett, M., Ferrer, E., & Ekstrom, A. D. (2020). How Much of What We Learn in
Virtual Reality Transfers to Real-World Navigation? *Multisensory Research*, *33*(4–5),
479–503. https://doi.org/10.1163/22134808-20201445

Hill, P. F., McAvan, A. S., Garren, J. D., Grilli, M. D., Barnes, C. A., & Ekstrom, A. D. (2023). *Age
differences in spatial memory are mitigated during naturalistic navigation* [Preprint].
Neuroscience. https://doi.org/10.1101/2023.01.23.525279

Ijaz, K., Ahmadpour, N., Naismith, S. L., & Calvo, R. A. (2019). An Immersive Virtual Reality
Platform for Assessing Spatial Navigation Memory in Predementia Screening: Feasibility and
Usability Study. *JMIR Mental Health*, *6*(9), e13887. https://doi.org/10.2196/13887

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness
Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International
Journal of Aviation Psychology*, *3*(3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3

Lakens, D. (2017). Equivalence Tests. *Social Psychological and Personality Science*, *8*(4), 355–362.
https://doi.org/10.1177/1948550617697177

Lakens, D., Scheel, A., & Isager, P. (2022). *Equivalence Testing for Psychological Research: A
Tutorial*. https://doi.org/10.1177/2515245918770963

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A Cross-Media Presence Questionnaire:
The ITC-Sense of Presence Inventory. *Presence: Teleoperators and Virtual Environments*,
10(3), 282–297. https://doi.org/10.1162/105474601300343612

Malinowski, J., & Horton, C. L. (2014). Evidence for the preferential incorporation of emotional
waking-life experiences into dreams. *Dreaming*, *24*(1), 18–31.
https://doi.org/10.1037/a0036017

Marraffino, M. D., Johnson, C. I., & Garibaldi, A. E. (2022). Virtual Reality is Better Than Desktop
for Training a Spatial Knowledge Task, but Not for Everyone. In J. Y. C. Chen & G.
Fragomeni (Eds.), *Virtual, Augmented and Mixed Reality: Design and Development* (Vol.
13317, pp. 212–223). Springer International Publishing.
https://doi.org/10.1007/978-3-031-05939-1_14

Martirosov, S., Bureš, M., & Zítka, T. (2022). Cyber sickness in low-immersive, semi-immersive, and

  fully immersive virtual reality. *Virtual Reality*, *26*(1), 15–32.

  https://doi.org/10.1007/s10055-021-00507-4

Mueller, S. C., Jackson, C. P. T., & Skelton, R. W. (2008). Sex differences in a virtual water maze: An

  eye tracking and pupillometry study. *Behavioural Brain Research*, *193*(2), 209–215.

  https://doi.org/10.1016/j.bbr.2008.05.017

Murcia-López, M., & Steed, A. (2016). The Effect of Environmental Features, Self-Avatar, and

  Immersion on Object Location Memory in Virtual Environments. *Frontiers in ICT*, *3*.

  https://doi.org/10.3389/fict.2016.00024

Murphy, M., Stickgold, R., Parr, M. E., Callahan, C., & Wamsley, E. J. (2018). Recurrence of

  task-related electroencephalographic activity during post-training quiet rest and sleep.

  Scientific Reports, *8*(1), 5398. https://doi.org/10.1038/s41598-018-23590-1

Nguyen, N. D., Tucker, M. A., Stickgold, R., & Wamsley, E. J. (2013). Overnight Sleep Enhances

  Hippocampus-Dependent Aspects of Spatial Memory. *Sleep*, *36*(7), 1051–1057.

  https://doi.org/10.5665/sleep.2808

Pan, X., & Hamilton, A. F. D. C. (2018). Why and how to use virtual reality to study human social

  interaction: The challenges of exploring a new research landscape. *British Journal of*

  Psychology, *109*(3), 395–417. https://doi.org/10.1111/bjop.12290

Piber, D., Nowacki, J., Mueller, S. C., Wingenfeld, K., & Otte, C. (2018). Sex effects on spatial

  learning but not on spatial memory retrieval in healthy young adults. *Behavioural Brain*

  Research, *336*, 44–50. https://doi.org/10.1016/j.bbr.2017.08.034

Picard-Deland, C., Nielsen, T., & Carr, M. (2021). Dreaming of the sleep lab. *PLOS ONE*, *16*(10),

  e0257738. https://doi.org/10.1371/journal.pone.0257738

Picard-Deland, C., Pastor, M., Solomonova, E., Paquette, T., & Nielsen, T. (2020). Flying dreams

  stimulated by an immersive virtual reality task. *Consciousness and Cognition*, *83*, 102958.

  https://doi.org/10.1016/j.concog.2020.102958

Saredakis, D., Szpak, A., Birckhead, B., Keage, H. A. D., Rizzo, A., & Loetscher, T. (2020). Factors

  Associated With Virtual Reality Sickness in Head-Mounted Displays: A Systematic Review

and Meta-Analysis. *Frontiers in Human Neuroscience*, *14*, 96.

https://doi.org/10.3389/fnhum.2020.00096

Schredl, M. (2006). *Factors Affecting the Continuity Between Waking and Dreaming: Emotional*

Intensity and Emotional Tone of the Waking-Life Event.

Simonsohn, U. (2015). Small Telescopes: Detectability and the Evaluation of Replication Results.

Psychological Science, *26*(5), 559–569. https://doi.org/10.1177/0956797614567341

Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing Our Lives with Immersive Virtual Reality.

Frontiers in Robotics and AI, *3*. https://doi.org/10.3389/frobt.2016.00074

Somarathna, R., Bednarz, T., & Mohammadi, G. (2022). Virtual Reality for Emotion Elicitation – A

Review. *IEEE Transactions on Affective Computing*, 1–21.

https://doi.org/10.1109/TAFFC.2022.3181053

Sousa Santos, B., Dias, P., Pimentel, A., Baggerman, J.-W., Ferreira, C., Silva, S., & Madeira, J.

(2009). Head-mounted display versus desktop for 3D navigation in virtual reality: A user

study. *Multimedia Tools and Applications*, *41*(1), 161–181.

https://doi.org/10.1007/s11042-008-0223-2

Spiers, H. J., Coutrot, A., & Hornberger, M. (2023). Explaining World-Wide Variation in Navigation

Ability from Millions of People: Citizen Science Project Sea Hero Quest. *Topics in Cognitive*

Science, *15*(1), 120–138. https://doi.org/10.1111/tops.12590

Srivastava, P., Rimzhim, A., Vijay, P., Singh, S., & Chandra, S. (2019). Desktop VR Is Better Than

Non-ambulatory HMD VR for Spatial Learning. *Frontiers in Robotics and AI*, *6*, 50.

https://doi.org/10.3389/frobt.2019.00050

Stamm, A. W., Nguyen, N. D., Seicol, B. J., Fagan, A., Oh, A., Drumm, M., Lundt, M., Stickgold, R.,

& Wamsley, E. J. (2014). Negative reinforcement impairs overnight memory consolidation.

Learning & Memory, *21*(11), 591–596. https://doi.org/10.1101/lm.035196.114

Thornberry, C., Cimadevilla, J. M., & Commins, S. (2021). Virtual Morris water maze: Opportunities

and challenges. *Reviews in the Neurosciences*, *32*(8), 887–903.

https://doi.org/10.1515/revneuro-2020-0149

Thorp, S., Sævild Ree, A., & Grassini, S. (2022). Temporal Development of Sense of Presence and

Cybersickness during an Immersive VR Experience. *Multimodal Technologies and Interaction*, *6*(5), 31. https://doi.org/10.3390/mti6050031

Tian, F., Hua, M., Zhang, W., Li, Y., & Yang, X. (2021). Emotional arousal in 2D versus 3D virtual reality environments. *PLOS ONE*, *16*(9), e0256211. https://doi.org/10.1371/journal.pone.0256211

Tian, F., Wang, X., Cheng, W., Lee, M., & Jin, Y. (2022). A Comparative Study on the Temporal Effects of 2D and VR Emotional Arousal. *Sensors*, *22*(21), 8491. https://doi.org/10.3390/s22218491

Wamsley, E. J., Hamilton, K., Graveline, Y., Manceor, S., & Parr, E. (2016). Test Expectation Enhances Memory Consolidation across Both Sleep and Wake. *PLOS ONE*, *11*(10), e0165141. https://doi.org/10.1371/journal.pone.0165141

Wamsley, E. J., & Stickgold, R. (2019). Dreaming of a learning task is associated with enhanced memory consolidation: Replication in an overnight sleep study. *Journal of Sleep Research*, *28*(1). https://doi.org/10.1111/jsr.12749

Wamsley, E. J., Tucker, M. A., Payne, J. D., & Stickgold, R. (2010). A brief nap is beneficial for human route-learning: The role of navigation experience and EEG spectral power. *Learning & Memory*, *17*(7), 332–336. https://doi.org/10.1101/lm.1828310

Wamsley, E. J., Tucker, M., Payne, J. D., Benavides, J. A., & Stickgold, R. (2010). Dreaming of a Learning Task Is Associated with Enhanced Sleep-Dependent Memory Consolidation. *Current Biology*, *20*(9), 850–855. https://doi.org/10.1016/j.cub.2010.03.027

Weech, S., Kenny, S., & Barnett-Cowan, M. (2019). Presence and Cybersickness in Virtual Reality Are Negatively Related: A Review. *Frontiers in Psychology*, *10*, 158. https://doi.org/10.3389/fpsyg.2019.00158

Weidner, F., Hoesch, A., Poeschl, S., & Broll, W. (2017). Comparing VR and non-VR driving simulations: An experimental user study. *2017 IEEE Virtual Reality (VR)*, 281–282. https://doi.org/10.1109/VR.2017.7892286

Wenk, N., Buetler, K. A., Penalver-Andres, J., Müri, R. M., & Marchal-Crespo, L. (2022). Naturalistic visualization of reaching movements using head-mounted displays improves movement

quality compared to conventional computer screens and proves high usability. *Journal of*

   NeuroEngineering and Rehabilitation, *19*(1), 137.

   https://doi.org/10.1186/s12984-022-01101-8

Woolley, D. G., Vermaercke, B., De Beeck, H. O., Wagemans, J., Gantois, I., D'Hooge, R., Swinnen,

   S. P., & Wenderoth, N. (2010). Sex differences in human virtual water maze performance:

   Novel measures reveal the relative contribution of directional responding and spatial

   knowledge. *Behavioural Brain Research*, *208*(2), 408–414.

   https://doi.org/10.1016/j.bbr.2009.12.019

Xie, J., Lan, P., Wang, S., Luo, Y., & Liu, G. (2023). Brain Activation Differences of Six Basic

   Emotions Between 2D Screen and Virtual Reality Modalities. *IEEE Transactions on Neural*

   Systems and Rehabilitation Engineering, *31*, 700–709.

   https://doi.org/10.1109/TNSRE.2022.3229389

Yao, R., Heath, T., Davies, A., Forsyth, T., Mitchell, N., & Hoberman, P. (2014). *Oculus VR Best*

   Practices Guide. http://developer.oculusvr.com/best-practices

Yu, C. K.-C. (2016a). Classification of typical dream themes and implications for dream

   interpretation. *Neuropsychoanalysis*, *18*(2), 133–146.

   https://doi.org/10.1080/15294145.2016.1236701

Yu, C. K.-C. (2016b). We dream typical themes every single night. *Dreaming*, *26*(4), 319–329.

   https://doi.org/10.1037/drm0000037

Zhao, J., Sensibaugh, T., Bodenheimer, B., McNamara, T. P., Nazareth, A., Newcombe, N., Minear,

   M., & Klippel, A. (2020). Desktop versus immersive virtual environments: Effects on spatial

   learning. *Spatial Cognition & Computation*, *20*(4), 328–363.

   https://doi.org/10.1080/13875868.2020.1817925