Final Paper

HCDE 516 - Experimental Research Methods
Wenbo Wang, Bokang Xu, Zhikang Xu

Introduction

Background

Virtual reality (VR) has gained popularity in video game experience. A lot of video games have adopted VR technology to create an immersive experience for the players. As the VR technology is still in its emerging stage, learning about how people navigate and memorize in a virtual environment (VE) setting is becoming an interesting field that can help game designers and developers to improve the gaming experience.

Previous research has been done on people's navigation and memorization behaviors in VE settings (Riva (Ph.D.) & Galimberti, 2001). However, little research has explored distracting factors on people's navigation experience in VE. In this paper, we aim to research the impact of static and dynamic visual distractions on people's navigation and memorization performance in VE setting.

Lit Review

Feasibility of Studying People's Navigation and Memory Performance in VE

VE is widely adopted by research in experimental psychology. Galimberti (2001) pointed out that the four main benefits of conducting an experimental study in VE, which includes: ecological validity, flexibility, sensorial feedback, and performance recording. According to previous research, navigation in VE is a feasible way to assess people's spatial memory and navigation performance (Riva (Ph.D.) & Galimberti, 2001). Faber (2015) suggests that assessing spatial memory in VR provides a fully controlled and easy-to-use experiment setting. Spatial memory performance has thus been measured in four different environments: real, virtual, virtual with directional information (compass), and hybrid. The most effective type of environment depends on the type of task provided (van der Ham, Faber, Venselaar, van Kreveld, & Löffler, 2015). van der Ham et. al found that both real and VE types of environment worked well for landmark and route knowledge, while a real environment invokes the best performance if survey knowledge is studied. The hybrid (mixing real and VE) condition was recommended when encountering survey knowledge and external restrictions.

Relationships between Memorization and Navigation Performance

Navigation involves basic human perceptual and memory-related processes (Taylor, Brunyé, & Taylor, 2008). There are two processes for people to navigate: local navigation and wayfinding (Eichenbaum, 2017). First, when people are in a currently perceived environment, they can directly and easily approach a directly observed location. Second, for the environment beyond the perceived space, people navigate utilizing recollection on learned routes or referring to an internalized cognitive map (Eichenbaum, 2017).

Additionally, people estimate the distance and absolute directions subconsciously when navigating around, and keep updating both egocentric (between self and objects) and allocentric (between objects) locations. Egocentric perception helps observers recognize the distance between the objects and themselves, while allocentric perception helps recognize the distance among objects. People maintain and constantly update allocentric locations based on internal and external spatial cues. When people navigate to an unseen target destination, they store their current point in their internalized allocentric maps and then compute the egocentric distance and directions of the goal (Zhao, 2016). From an egocentric perspective, the degree of distance and the visibility of a landmark influence people's perception of a landmark (Karimpur, Röser, & Hamburger, 2016). Therefore, people's perception of allocentric and egocentric locations may influence their navigation performance.

There are potential relationships between navigation and memorization performance in mazes. In mazes, navigating to the wrong route is a frequent reality. Hence, memorizing the past routes is highly likely to help decrease the maze navigation time (Ballard, Hayhoe, & Pelz, 1995, Flanagan & Johansson, 2003, Kong, Schunn, & Wallstrom, 2010). In addition, a sense of direction is also important when navigating in mazes. Lacking such quality are more likely to get lost and results in worse navigation performance (Wollan, 2018).

External Factors that Influence People's Navigation Performance

A real-world navigation system usually depends on the quality of the mechanical system (such as sensors and actuators), the complexity of the environment, and the corresponding models (Fulgenzi, 2009). In navigation context, static and changing environment (such as paths containing pedestrians, human-driven vehicles, and robots passing-by) contribute to different reactions in navigation experience, and an unknown dynamic environment is more restraining in terms of real-time decision and time of reaction than an unknown static environment (Fulgenzi, 2009).

Approaches to Assess People's Navigation and Memorization Performance in VE

Some of the common measures for assessing human navigation abilities in VE are landmark recognition, route distance in the proportion of actual distance, route position, and pointing map drawing. For example, (van der Ham et al., 2015) utilized these measurements to study different navigation performance under different environments, by asking participants to navigate from the start to the destination and providing a post-task survey. Parush & Berman (2004) created further detailed measures including navigation duration, number of navigation steps, unessential navigation steps, target correctness, orientation correctness (the number of times the participant correctly pointed in the direction of the required target), orientation response time, and orientation deviation. Their approach took a closer investigation on the navigation detail data and analyze how they relate to different conditions.

Background

All studies mentioned above have validated VE navigation experiments as a valuable way of assessing memory and navigation performance, revealed potential relationships among variables, and provided guidance for measurements. Yet little research addresses navigation distractions in VE settings, and few field tests were conducted to validate relevant hypotheses. Therefore our research intends to cover the gap between existing studies and the impact of distractions in VE navigation and to make recommendations to VR designers and developers based on the results of our study.

Hypotheses

We hypothesize that moving states of visual distractions have a significant impact on people's navigation and memory performance in VE such that dynamic visual distractions will result in poorer navigation performance when compared to static. We also hypothesize this will result in poorer memory performance for dynamic visual distractions when compared to static visual distractions in VE.

Methods

Study Tool

We chose to use Minecraft VR to set up mazes in VE. Minecraft is a classic adventure game which enables people to navigate and explore mazes. The VR version was released in 2016. The latest Minecraft VR game contains a total of 677 static and dynamic native building blocks ("Block – Official Minecraft Wiki," n.d.). It allows users to create various customized mazes easily without heavy coding.

Variables

The independent variable (IV) of our study is the moving state of the visual distractions. The dependent variables (DVs) are people's navigation and memorization performance. The control variables (CVs) include VE, customized mazes in Minecraft VR and non-hardcore Minecraft VR players.

Experiment Design

This experiment adopts a pretest and posttest design.

Treatment Groups

Participants are randomly assigned into two groups, one encounter only dynamic visual distractions (zombie horses, sheep, pigs, and chickens) in a customized maze in Minecraft VR, while the other encounter only static ones (various types of flowers, grass) in the maze with exactly the same route.

Tutorial

Before the real maze, participants are asked to complete a short tutorial maze to get familiarized with Minecraft VR controls and the maze environment. The tutorial maze has the same setting as the real maze, so we can make sure the participants share the same capability to complete the task. They also learn that the distractions do not actually block their navigation route. After the tutorial maze, participants will enter the real maze.

Main Task

Participants are informed that their sole task is to navigate to the end of the maze which is a room with an 'End of the Maze' billboard as quickly as possible. We stage the experiment by not telling the participants that their spatial memory would also be assessed in the posttest, to make sure they do not wander about or spend additional time memorizing anything intentionally. The timer starts when the participants start at the beginning of the actual maze marked by a 'Start of the Maze' billboard.

Data Collection

We collect three types of data during the test: completion time, landmark recall test score, and point-to-the-end test score.

Completion Time

Completion time is defined as the time the participants use to finish the real maze.

Landmark Recall Test Score

The landmark recall test score is collected by the posttest questionnaire. We ask participants to select landmarks they remember seeing in the maze. We set up six distinctive landmarks on the critical path, and four of them are included in the questionnaire among other confusing options. Both groups are given the same number of correct answers, and we count the number of correct answers as the final test score.

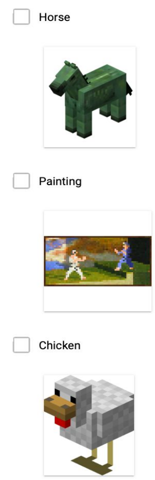

Figure 1. Landmark Recall Test Screenshot

Point-to-the-end Test Score

The point-to-the-end test score is also collected in the posttest questionnaire. We ask the participants to point out the direction of the final destination from four different options, assuming that they are standing in front of the entrance.

Figure 2a. Point-to-the-end baseline

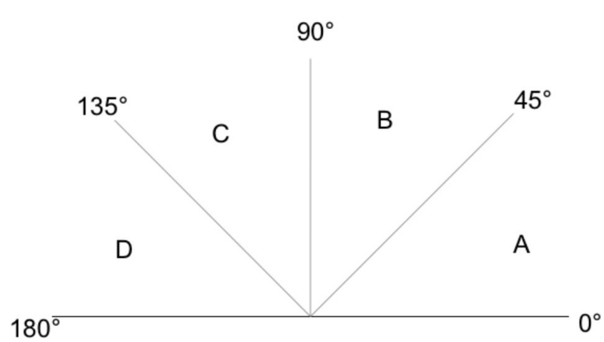

Figure 2b. Point-to-the-end Test Screenshot

Actual Study

Pretest / Screening Process

We recruited participants by posting an online survey on social media platforms. The online survey contains direct inquiries to learn about participants' past VR / gaming experience. We used the survey to measure participants' navigation and visual memory capabilities and screened out participants with extensive experience with Minecraft VR since they are highly likely to be aware of relying solely on minor color differences to navigate through Minecraft mazes and to ignore and bypass all distractions too easily. We also screen out participants with extraordinary navigation ability, memory capability, and maze solving ability by leveraging online spatial and visual memory tests (link: https://www.psytoolkit.org/experiment-library/experiment_corsi.html and https://www.humanbenchmark.com/tests/memory). In the same online survey, we also collected participants' age and gender to reveal potential confounding variables.

Pilot-test

We ran a pilot-test with the first three participants. However, they didn't go through the maze due to motion sickness, so we discovered the severity of motion sickness may be a potential confounding variable. We also found that the original maze size may cause a high dropout rate, hence, we reduced the maze size to ¼ of the original one.

Participants

After excluding the first three participants in the pilot-test, we have 16 participants left to finish testing on the maze. Their ages range from 21 to 54 and we have seven male and nine female. None of them has had extensive VR or Minecraft experience.

Table 1a. Participants Gender Distribution

Gender	Count of Participants
Male	7
Female	9

Table 1b. Participants Age Distribution

Age Range	Count of participants
21 to 36	15
36 to 54	1

Results

Analysis of the effects between Independent Variable and Dependent Variable

Statistical Test Methods

The statistical test methods are listed in the table below. The motion states of the visual distractions are a categorical and nominal variable. The maze completion time, measured in seconds, should be considered as a continuous ratio variable. Thus the statistic test method we choose is t-test. The second dependent variable is the landmarks recall test score, which has five possible values: 0, 1, 2, 3 and 4. It is considered as a categorical variable and thus the statistic test method we choose is Chi-squared test. The third dependent variable is the point-to-the-end test score, which has two possible values: right and wrong. It is considered as a categorical variable and thus the statistic test method we choose is Chi-squared test.

Table 2. Statistical Test Methods

	DV1: Maze Completion Time (Ratio+Continuous)	DV2: Landmarks Recall Test score (Categorical)	DV3: Point-to-the-end test score (Categorical)
IV: Dynamic / Static Visual Distractions (Categorical)	t-test	Chi-squared test	Chi-squared test

Statistical Test Results

With α=0.05, there is no significant difference between the impact of dynamic and static distractions on navigation performance. Nor is there such difference between the impact of dynamic and static distractions on spatial memory performance, or between the impact of dynamic and static distractions on visual memory performance. However, with α=0.1, there is a significant difference between the impact of dynamic and static distractions on visual memory performance.

In a nutshell, we do not find a statistical difference between our two groups, and therefore we fail to reject the null hypothesis. However, with α=0.1, such difference exists between the impact of dynamic and static distractions on visual memory performance.

Table 3. Statistical Test Results

	IV~DV1: Maze Completion Time (Ratio+Continuous)	IV~DV2: Landmarks Recall Test Score (Categorical)	IV~DV3: Point-to-the-end Test Score (Categorical)
Method	Unpaired two-samples t-test	Chi-squared test	Chi-squared test
p-value	0.461	0.08894	0.8087
Significant?	N	N	N

Statistical Test Details

The Effects of Dynamic and Static Visual Distractions on Maze Completion Time

Data Exploration and Visualization

Table 4. Data Exploration of the Maze Completion Time by groups

Group	Count	Mean	sd
1	9	99.78	55.83
2	7	77.00	64.34

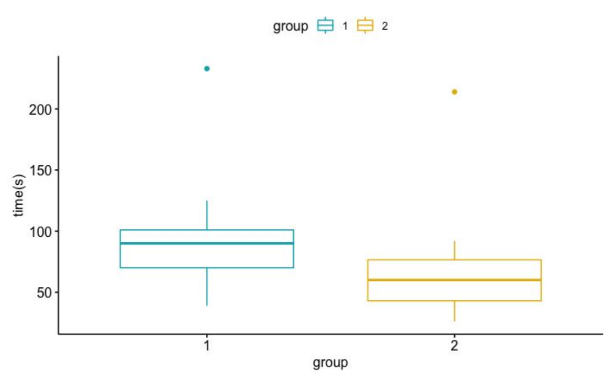

Figure 3. Boxplots for groups

Assumption Examination

First, the two groups are independent since the samples from group 1 and group 2 are not related.
Second, we used the Shapiro-Wilk normality test to find out whether the data from the two groups is normally distributed. The null hypothesis of the test is that the data is normally distributed, while the alternative hypothesis being the data is not normally distributed. We use functions with() and shapiro.test() to conduct the Shapiro-Wilk test for each group of samples.

Code and result:

```
with(data, shapiro.test(time[group == "1"]))# p =0.28
with(data, shapiro.test(time[group == "2"]))# p =0.13
```

From the output, we discover that the two p-values are greater than the significance level 0.05 implying that the distribution of the data is not significantly different from the normal distribution. In other words, we can assume normality.
Third, we investigate that if the two populations have the same variances. We use F-test to test for homogeneity in variances. This is performed with the function var.test() as follows:

Code:

```
var.test(time ~ group, data = data)
```

Result:

```
F-test to compare two variances
data:  time by group
F = 0.75282, num df = 8, denom df = 6, p-value = 0.691
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.134442 3.501909
sample estimates:
ratio of variances
         0.7528243
```

The p-value of F-test is $p = 0.691$. It's greater than the significance level of $\alpha = 0.05$. In conclusion, there is no significant difference between the variances of the two sets of data. Therefore, we can use the classic t-test which assumes equality of the two variances.

Compute unpaired two-samples t-test

Code:

```
t.test(time ~ group, data = data, var.equal = TRUE)
```

Result:

```
Two Sample t-test

data:  time by group
t = 0.75797, df = 14, p-value = 0.461
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -41.67536  87.23091
sample estimates:
mean in group 1 mean in group 2
       99.77778        77.00000
```

The p-value of the test is 0.461, which is greater than the significance level α = 0.05. We can conclude that group 1's maze completion time is not significantly different from group 2's maze completion time.

The Effects of Dynamic/Static Visual Distractions on Landmarks Recall Test Score

Data Exploration and Visualization

Table 5. Data Exploration of the Landmarks Recall Test Score by groups

	Group 1	Group 2
1 correct answer	3	3
2 correct answers	5	1
3 correct answers	0	3
4 correct answers	1	0

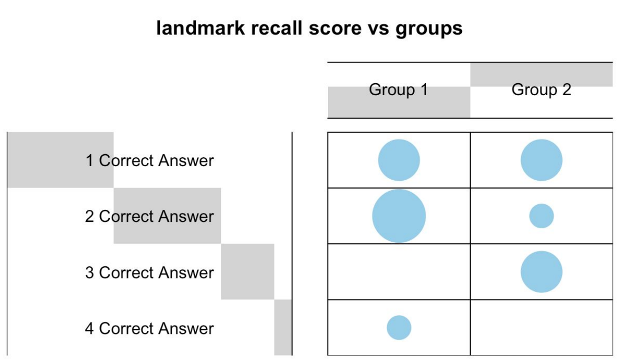

Figure 4. Landmark result-test score distribution

Hypothesis

Chi-square test examines whether rows and columns of a contingency table are statistically significantly associated. The null hypothesis (H0) is that the row and column variables of the contingency table are independent while the alternative hypothesis (H1) being the row and column variables are dependent.

Code and Result

```
tbl = table(data3$score_land, data3$group)
colnames(tbl) <- c("Group 1", "Group 2")
rownames(tbl) <- c("1 Correct Answer", "2 Correct Answer", "3 Correct Answer", "4
Correct Answer")
chisq.test(tbl)

Pearson's Chi-squared test
data:  tbl
X-squared = 6.5185, df = 3, p-value = 0.08894
```

Conclusion

From the output, the row and the column variables are not statistically significantly associated (p-value = 0.08894).

The Effects of Dynamic/Static Visual Distractions on Point-to-the-end Test Score

Data Exploration and Visualization

Table 6. Data Exploration of the Point-to-the-end Test Score by groups

	Group 1	Group 2
correct	8	5
wrong	1	2

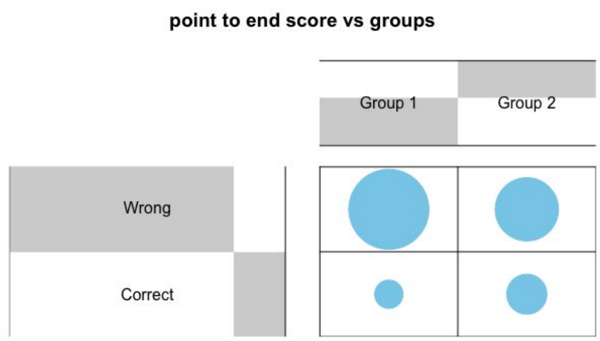

Figure 5. Point-to-the-end score distribution

Hypothesis

Chi-square test examines whether rows and columns of a contingency table are statistically significantly associated. The null hypothesis (H0) is that the row and column variables of the contingency table are independent and the alternative hypothesis (H1) is that the row and column variables are dependent.

Code and Result

```
tbl = table(data3$score_point, data3$group)
colnames(tbl) <- c("Group 1", "Group 2")
rownames(tbl) <- c("Wrong", "Correct")
chisq.test(tbl)

Pearson's Chi-squared test with Yates' continuity correction

data:  tbl
X-squared = 0.058608, df = 1, p-value = 0.8087
```

Conclusion

From the output, the row and the column variables are not statistically significantly associated (p-value = 0.8087).

Statistical Test of Potential Confounding Effects

Statistical Test Methods

We then explored the impact of potential confounding variables on navigation performance, spatial memory performance, and visual memory performance. We collected the following as potential confounding variables: spatial memory test score, visual memory test score, gender, previous video game experience, previous VR experience, and previous Minecraft experience.

Table 7. Statistical Test Methods of the Potential Confounding Effects

	V1: Maze Completion Time (Ratio+Continuous)	V2: Landmarks Recall Test Score (Categorical)	V3: Point-to-the-end Test Score (Categorical)
Spatial Memory Test Score (Ratio+Continuous)	Pearson's correlation test	Kruskal-Wallis one-way ANOVA test	t-test
Visual Memory Test Score (Ratio+Continuous)	Pearson's correlation test	Kruskal-Wallis one-way ANOVA test	t-test
Gender (Categorical)	t-test	Chi-squared test	Chi-squared test
Previous Video Game Experience(Categorical)	t-test	Chi-squared test	Chi-squared test
Previous VR Experience(Categorical)	t-test	Chi-squared test	Chi-squared test
Previous Minecraft Experience(Categorical)	t-test	Chi-squared test	Chi-squared test

Statistical Test Result

Table 8. Statistical Test Results(p-value) of the Potential Confounding Effects

	V1: Maze Completion Time (Ratio+Continuous)	V2: Landmarks Recall Test Score (Categorical)	V3: Point-to-the-end Test Score (Categorical)
Spatial Memory Test Score (Ratio+Continuous)	0.0225	0.7651	0.0918
Visual Memory Test Score (Ratio+Continuous)	0.0727	0.9435	0.1775
Gender (Categorical)	0.9224	0.6198	0.2942
Previous Video Game Experience(Categorical)	0.764	0.6198	0.0014
Previous VR Experience(Categorical)	0.9034	0.7851	0.0688
Previous Minecraft Experience(Categorical)	0.7379	0.6198	0.9183

With α=0.05, we found two potential confounding effects: spatial memory test score may have a confounding effect on navigation performance; previous video game experience is highly likely to have a confounding effect on spatial memory performance. With α=0.1, we found several additional potential confounding effect: previous VR experience may have a confounding effect on spatial memory performance; visual memory test score may have a confounding effect on navigation performance; spatial memory test score may have a confounding effect on spatial memory performance

Power Analysis

In each group, to obtain a power of 0.80, when the effect size is moderate (0.3) and a significance level of 0.05 is employed, the corresponding recommended sample sizes are listed in the table below.

Table 9. Power Analysis

	DV1: Maze Completion Time	DV2: Landmarks Recall Test Score	DV3: Point-to-the-end Test Score
IV: Dynamic/Static Visual Distractions	96	44	121

Discussion

Conclusion

According to the results shared earlier, the conclusion we draw is that no statistically significant difference is found between static and dynamic distractions in their impacts on people's navigation and memorization performance. Considering our small sample size, we also examined the scenario where $\alpha=0.1$. In this condition, we found a significant difference between the impact of dynamic and static distractions on visual memory performance. We recommend a larger sample size to examine this relationship.

Confounding Effects

We discovered that previous video game experience has a confounding effect on spatial memory performance, and spatial memory test score has a confounding effect on navigation performance.

This paper finds it reasonable since people with richer past video game experience may have more exposure to navigation games and thus gained more spatial memory practice, which could result in their better performance in the spatial memory test in the study. Additionally, as our aforementioned literature has pointed out, memorizing the past routes is highly likely to help decrease the maze navigation time (Ballard, Hayhoe, & Pelz, 1995, Flanagan & Johansson, 2003, Kong, Schunn, & Wallstrom, 2010). This validates our conclusion that people with better spatial memory capabilities are more likely to perform better in navigation tasks.

Considering our small sample size, we also examined the potential confounding effects in the scenario where $\alpha=0.1$. In this condition, we discovered that previous VR experience has a confounding effect on spatial memory performance. We also found that visual memory test score has a confounding effect on navigation performance. Last but not least, we discovered that the spatial memory test score has a confounding effect on spatial memory performance. We recommend a larger sample size to examine such effects.

Limitations

In reflection, we were limited by the time and funding for recruitment, hence our sample size is too small. According to Bohannon(2015), the minimum sample size for a group of a study to be considered seriously by academia should be 30. Furthermore, we screened out people with extensive Minecraft or VR experience or extraordinary memorization capabilities, hence our study may lose the potential to discover interaction effect among these factors. We also had to reduce the size of our maze, since current VR technology introduced significant motion sickness and resulted in quite a few participants' failures to complete the task. This reduction of map size and complexity may have introduced more noise to our final results and impacted our ability to reject the null hypothesis, as shorter and simpler mazes may provide less time for people to adjust to the environment and increase confounding effects of the environment.

Recommendations

We propose two suggestions for VR game designers and developers. First, reduction in users' motion sickness is vital for further study in VE settings as well as to increase the adoption of VR technology. This could be potentially achieved by improving clarity and resolution of the VR visuals. Second, even though our experiment does not reject the null hypothesis, we think our experiment design provides a valuable way of measuring qualities of distractions and may help decisions on placement strategies of gaming objects, such as NPC or treasure. We recommend further research in a similar fashion to examine our hypotheses.

For future studies, we recommend recruiting more participants. We also recommend waiting for advancement in VR technology to reduce motion sickness significantly and then conduct similar experiments with mazes of more diverse complexity to potentially discover greater differences. Another possible improvement is to arrange denser visual distractions to attain more delay on distractions. To create better equality between the two groups, consider controlling more confounding variables, such as motion sickness as well as spatial and visual memory capabilities.

Appendix. Reference

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, *7*(1), 66–80. https://doi.org/10.1162/jocn.1995.7.1.66

Block – Official Minecraft Wiki. (n.d.). Retrieved May 16, 2019, from https://minecraft.gamepedia.com/Block

Eichenbaum, H. (2017). The role of the hippocampus in navigation is memory. *Journal of Neurophysiology*, *117*(4), 1785–1796. https://doi.org/10.1152/jn.00005.2017

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*(6950), 769–771. https://doi.org/10.1038/nature01861

Fulgenzi, C. (2009). *Autonomous navigation in dynamic uncertain environment using probabilistic models of perception and collision risk prediction.* 141.

Karimpur, H., Röser, F., & Hamburger, K. (2016). Finding the Return Path: Landmark Position Effects and the Influence of Perspective. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01956

Kong, X., Schunn, C. D., & Wallstrom, G. L. (2010). High regularities in eye-movement patterns reveal the dynamics of the visual working memory allocation mechanism. *Cognitive Science*, *34*(2), 322–337. https://doi.org/10.1111/j.1551-6709.2009.01075.x

Riva (Ph.D.), G., & Galimberti, C. (2001). *Towards Cyberpsychology: Mind, Cognition, and Society in the Internet Age*. IOS Press.

Taylor, H. A., Brunyé, T. T., & Taylor, S. T. (2008). Spatial Mental Representation: Implications for Navigation System Design. *Reviews of Human Factors and Ergonomics*, *4*(1), 1–40. https://doi.org/10.1518/155723408X342835

van der Ham, I. J. M., Faber, A. M. E., Venselaar, M., van Kreveld, M. J., & Löffler, M. (2015).

Ecological validity of virtual environments to assess human navigation ability. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00637

Wollan, M. (2018, October 18). How to Navigate a Maze. *The New York Times*. Retrieved from https://www.nytimes.com/2018/10/17/magazine/how-to-navigate-a-maze.html

Zhao, C. (2016). *Understanding Human Spatial Navigation Behaviors: A Cognitive Modeling Approach*. Retrieved from https://etda.libraries.psu.edu/catalog/28590