3D Geographic Model VS Street View Panorama

a Cognitive Study on Navigation in Different Google Maps Representations

by
Tianxia Gu

Submitted to the Department of Architecture
in Partial Fulfillment of the Requirements for the Degree of

Bachelor of Science in Architecture Studies

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2018

Signature redacted

Signature of Author ....
Department of Architecture
May 24, 2018

Signature redacted

Certified by ....
Takehiko Nagakura
Associate Professor of Design and Computation
Thesis Supervisor

Signature redacted

Accepted by ....
Leslie Keith Norford
Professor of Building Technology
Department of Architecture Undergraduate Officer

3D Geographic Model VS Street View Panorama

a Cognitive Study on Navigation in Different Google Maps Representations

by
Tianxia Gu

Submitted to the Department of Architecture
on May 24, 2018 in Partial Fulfillment of the Requirements for
the Degree Bachelor of Science in Architecture Studies at Massachusetts Institute of Technology

Abstract:

Throughout history, different tools have been invented to help people navigate in space. The different tools use different modes of representation as abstractions of 3-dimensional space. Two of the representations used in Google Maps, a modern wayfinding technology, are the 3D geographic model mode and the street view panorama mode. In this thesis, we explore the wayfinding behaviors of people in those two representations by performing experiments. We find that each representation mode is advantageous for a different type of city structure (e.g., regular grid vs. irregular streets). Each representation mode is also preferred by people according to their spatial processing type preference and the wayfinding task type they perform. After evaluating our findings from the experiments, we propose a design of a new representation with facade images augmenting a 2D satellite map. We believe this design incorporates the advantages of both representations studied.

Thesis Supervisor: Takehiko Nagakura
Title: Associate Professor of Design and Computation

Acknowledgements

To the merciful lord, for all the countless gifts you offered.

To my parents, for all your unconditional love, care, and support throughout my life journey.

To the participants in my experiments, for your interest in my research topic and your precious time and energy to help me collect meaningful data for my research.

To the members of MIT Design and Computation Group, for your vital feedback to help me further my research.

To all my dearest friends all over the world, who cheered, laughed, and smiled with me. You all enriched my life.

To Run Chen, Jun Jie Joseph Kuan, Chung-Yueh Lin, Zilu Pan, Chengkai Zhang, and Yunkun Zhou, for being my best friends at MIT, brightening even the dimmest of days and being there for me to lean on.

To Renee Caso, who kept an eye on all of us architecture undergraduates and supported us unconditionally.

To my major advisors Terry Knight and Bruce Tidor, and my minor advisor Ben Ross Schneider, who supported me throughout my years at MIT and always believed in me, even when I took the road less traveled by.

Finally, my eternal thanks to Takehiko Nagakura, my thesis advisor, for your scintillating conversations and valuable insights. It has been a pleasure working with you.

Table of Contents

Introduction

Wayfinding encompasses all of the ways in which people orient themselves in physical space and navigate from place to place. In the context of architecture, it refers to the user experience of orienting and choosing a path within the built environment. Wayfinding has been an existent problem in history, and different tools have been invented to help people navigate. Examples of such tools include maps, which provide the plan of the area for the user to identify different locations and paths between them, as well as modern software such as Google Maps, which has both up-to-date geodata and a built-in routing algorithm to come up with possible paths. However, both conventional maps (Fig. 1a) and Google Maps user interface (UI) (Fig. 1b) rely on abstracted 2-dimensional data. While 2-dimensional data might be a good method of presenting geo-information within a specified area, 2-D data are highly abstracted and not very intuitive for novices since they miss specific information, such as the heights and the façade views of the buildings, which are sometimes important for people to identify architecture and find ways within an area.

a

b

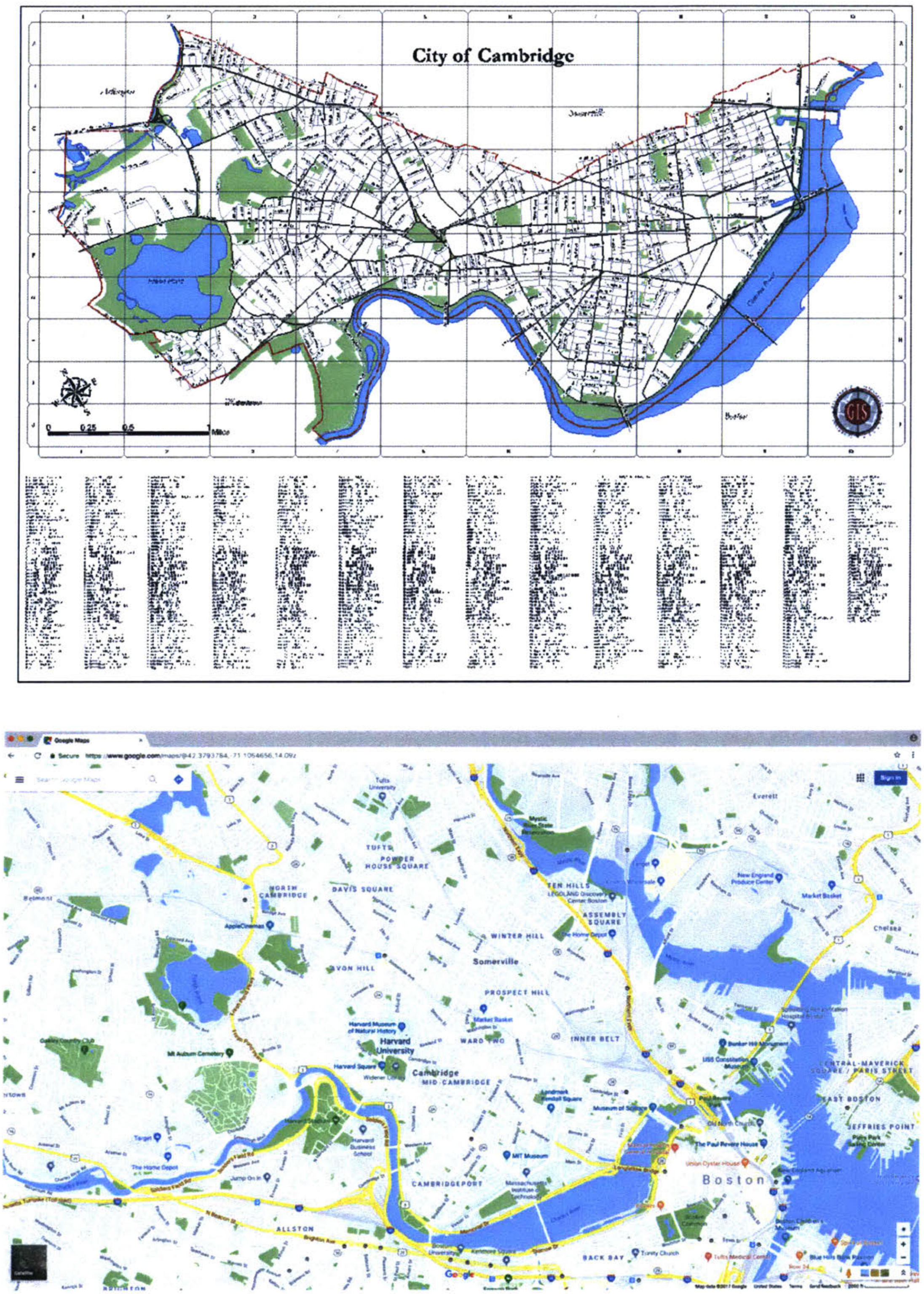


Figure 1. 2-dimensional map representations of Cambridge, MA, USA in (a) physical map, from:"Cambridge USA." *Visit Cambridge Massachusetts | Cambridge Office of Tourism*. Accessed 23 Nov. 2017; and (b) Google Maps, from: "Cambridge, Massachusetts." Map, *Google Maps*. Accessed 20 Nov. 2017.

A modern approach to the wayfinding problem involves the use of 3-dimensional media, including 3D geographic models, panoramas, and virtual reality (VR). 3D geographic models (Fig. 2a) are usually retrieved by stitching together images captured using drones. 3D geographic models have additional information about the heights of the architectures and landscapes, and are more realistic compared to the conventional 2D maps; panorama (Fig. 2b) is a 360-degree visual medium that allows the user to experience a 360-degree representation of a place. Compared to 2D photos and videos, panoramic photos and videos contain more information about the surroundings of a specific place; VR (Fig. 2c) is a computer technology that is used to generate realistic images, sounds and other sensations that simulate a user's physical presence in a virtual environment. VR places the user in an immersive environment that is very similar to the physical space. Therefore, wayfinding using 3-dimensional media provides more realistic spatial information for inexperienced users, and is thus more intuitive than conventional wayfinding methods.

a

b

c

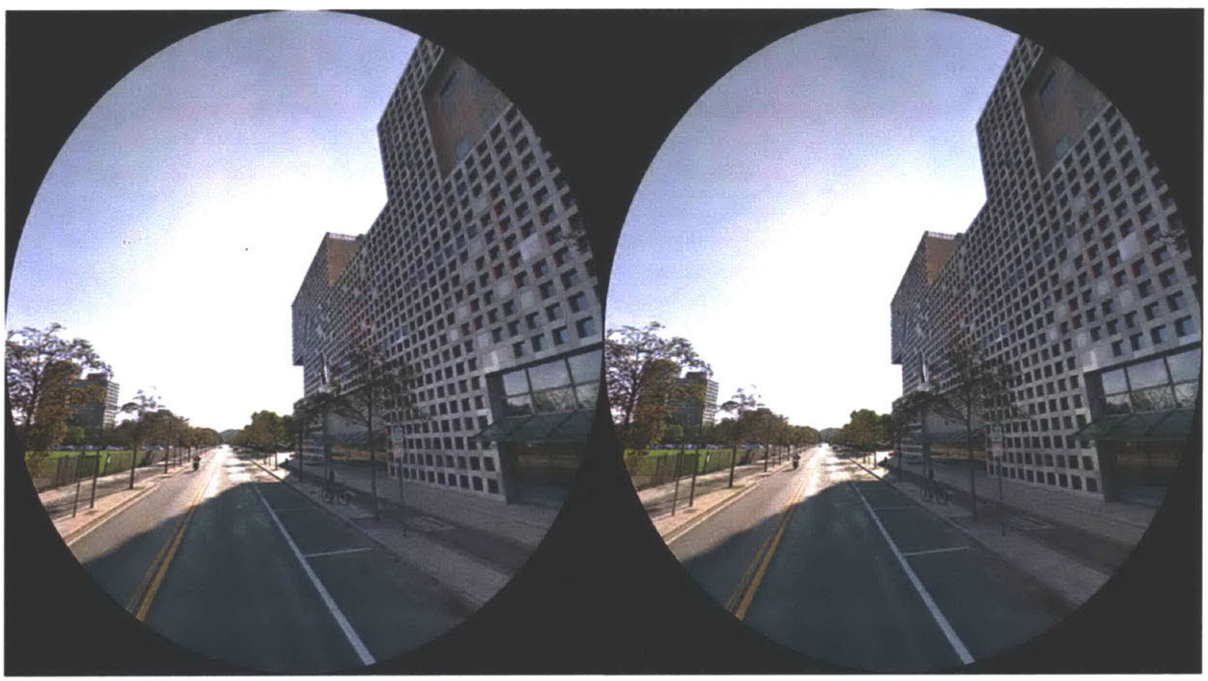


Figure 2. Three different representations of the location in front of MIT Simmons Hall on Vassar Street, from: "Cambridge, Massachusetts." Map, *Google Maps*. Accessed 23 Nov. 2017. (a) is the 3D mode of Google Maps with 3D models; (b) is the street view mode of Google Maps with street view panorama; (c) is each eye's view of the 3D panoramic view in VR.

However, without including a 2-dimensional abstracted plan, wayfinding in 3-dimensional media can sometimes be difficult because users have a difficult time perceiving the space as a whole, and/or identifying different locations within the space and the paths between them. In order to get detailed information of the architectures and landscapes from the 3-dimensional media, the user usually needs to zoom in and look very closely, losing information on spatial organization and orientation of architectures.

In order to solve some of the deficiencies of 2-D and incorporate the benefits of 3-D, efforts are made to incorporate both an immersive 3-dimensional experience and 2-dimensional abstracted geo-information in wayfinding software. One example is Google Earth VR, which allows people to both take a bird's eye view of the space and also to put themselves in the space and walk around. Another example is VR video that has augmentations of 2-dimensional maps embedded to provide the users with more spatial information. In addition, the new Google Maps UI also has different modes that allows the user to toggle between different representations: the 3D mode (Fig. 2a) allows the user to view the 3D geographic models of high resolution in some cities at a specific perspective angle to the 2D map representation, and the street view mode (Fig. 2b) allows the user to view the panoramic photos on different streets and locations alongside a zoomed-in 2D map representation.

While incorporating both a 3-dimensional experience and 2-dimensional abstracted geo-information in wayfinding, it is important to show the correlation between the two types of information. It is also important to study how the two types of information correlate in different representations, and how the different representations are effective or not in terms of in wayfinding for users. Since Google Maps has multiple representations to help people navigate, this thesis investigates the strengths and weaknesses of both the 3D mode and the street view mode, and proposes a framework for an optimal representation that incorporates the strengths of all the existing methods.

To achieve my goal, I plan to conduct cognitive wayfinding experiments on participants at MIT at a variety of locations using the two representations. I will collect both quantitative data on the wayfinding performance of the participants, and qualitative data on the UX feedback of the participants on the tools. I will analyze the data to draw conclusions and propose a framework of an optimal method using my conclusions. If possible, I would also like to design and develop a software as a proof of concept of my framework.

Literature Review

Since the research involves both the investigation of human's cognitive behavior of wayfinding and the evaluation of Google Maps' UI representations, literature in wayfinding, UI design, as well as their connections is reviewed below in three sections.

Wayfinding and Cognitive Mapping

The concept of wayfinding is first defined in Kevin Lynch's seminal work, the Image *of the City*. In the book, Lynch defines *wayfinding* as "a consistent use and organization of definite sensory cues from the external environment" (3). Lynch further defines *wayfinding* as a cognitive behavior. Following the idea of cognitive mapping and its relations to wayfinding, further research suggests that in order to perform wayfinding, humans "acquire, code, store, decode and use cognitive information as part of their navigation and wayfinding activities" (Golledge, ch. 1). From the cognitive perspective, wayfinding is "the process of determining and following a path or route between an origin and a destination" (Golledge, ch. 1). In order to perform wayfinding tasks, information is "extracted from large-scale external environments and stored in human memory," and "exists in some type of psychological space whose metricity may be unknown" (Golledge, ch. 1). Therefore, since different people have different capabilities in extracting and storing information, different people also have different performance on wayfinding tasks, known as as spatial abilities.

Past research evaluating spatial abilities take different approaches: some researchers evaluate spatial abilities using some psychometrics, including information-processing and developmental traditions such as *Visualization and Mental Rotation*, *Visual-Spatial Memory*, *Mental Imagery*, *Spatial Perspective-Taking and Orientation*, *Map Interpretation and Generalizations* (Golledge, ch. 2). Those can be tested by incorporating tasks resembling items on psychometric tests in experiments to "differentiate fundamental processes on the basis of the time necessary to respond accurately and, in some instances, error analysis" (Golledge, ch. 2); some researchers believe in evaluating the behavioral level of wayfinding, the making of spatial choice (Golledge, ch. 3). In this case, they argue that the evaluation should focus on spatial attributes essential to spatial choice, such as *distance* and *spatial configuration*, and to model more realistic wayfinding situations, should also take into account trade-offs against non-spatial attributes such as *time* and *priority* (Golledge, ch. 3); other researchers argue that since the goal of wayfinding is to reach the target destination from the start point, the evaluation of wayfinding should focus on the evaluation of two most important processes involved in achieving the goal: *piloting* and *path-integration* (Golledge, ch. 5). *Piloting* is position-based navigation that "relies on external signals indicating the traveler's position," and *path-integration* is velocity-based navigation that "relies on external or internal signals indicating the traveler's course and speed" (Golledge, ch. 5). In addition to path-finding between two places, real-life wayfinding also involves cases of self-locating based on landmarks. Therefore, researchers also propose that another important aspect of wayfinding evaluation is *landmark-space association*, which "helps the subject to return to particular locations on a long-term basis" (Golledge, ch. 8). While information extracted and stored for wayfinding exists in some type of psychological space whose metricity may be unknown, past efforts have been made to evaluate people's spatial abilities, which to some extent reflect their information-processing abilities.

User Interfaces for Space Representations and Wayfinding Tools

Tools have been developed throughout history for wayfinding, including maps, written descriptions, and various forms of image representations, all of which help humans make spatial decisions and guide their movement behavior. Thus, it is implied in the research that in order for the tools for wayfinding to be effective, they ought to aid human cognition in a specific way that makes it easy for humans to recognize, remember and navigate.

The most common techniques in earlier times involved representation and abstraction of the physical space, in forms of text description, voice guide and map images; modern technologies including 3D media and virtual reality add to the inventory of such representation and abstraction. Past research has evaluated the effectiveness in different representation and abstraction schemes of spaces, and on comparing and contrasting conventional representation and 3D virtual reality representation in their effectiveness in mimicking the physical space and allowing people to perform certain tasks in the space (Heydarian et al., 116-126; Choi, 1-58; Bliss, Tidwell and Guest, 73-86). Many of the research projects use cognitive experiments and analyze both quantitative performance data and qualitative data to reach conclusions, in alignment with Lynch's theory of cognitive mapping.

A 2015 paper by Heydarian et al. evaluates whether immersive virtual environments (IVEs) are adequate representations of physical environments, specifically office environments, by examining the difference in end-users' performances in office-related activities within a physical office space and a virtual office space (118-120). Heydarian et al. explore the use of IVE for an office space, evaluate the end-user's sense of presence within the IVE through a questionnaire, and compare user performance on a set of identical tasks in an IVE and a physical environment with same architectural settings (120-122). Based on the performance measures and questionnaire data, Heydarian et al. conclude that IVE is a satisfactory representation of the physical environment (122-125).

A 2016 architecture thesis by Choi compares and contrasts multiple frameworks in space representation, in terms of scale, proportion, orientation and location of objects in the space (ch. 3). The results showed that Projected Digital Space (PDS) and Immersive Virtual Space (IVS) have different strengths in representing the physical space: PDS refers to the space mapped out and projected on flat 2-dimensional screen. Its strength is in simplifying a space with objects on the flat 2D screen which enables users to perceive the space as a whole, giving an opportunity for pattern recognition which is suitable when studying spatial organization and orientation of different objects; IVS refers to virtual 3-dimensional immersive space that can be experienced by wearing a virtual reality headset. Its strength is in simulating visual perception of physical space with high accuracy (ch. 3).

In terms of wayfinding, there has been research early in the 90s comparing the effectiveness of the use of virtual reality and the use of schematic building blueprints and verbal directive description in training firefighters in their spatial navigation (Bliss, Tidwell and Guest, 77-80). While there are no significant differences in the speed and accuracy of the firefighters trained using the two methods in their search, the experiment had a few limitations (81-85). First, the navigation is too specific to be representative of general wayfinding behavior of humans; also, the virtual reality technologies have developed throughout the past few years and have become much more proficient in spatial representation, compared to those in the 90s. Thus, wayfinding is worth revisiting as a comparative study of different schemes in spatial representation to evaluate the cognitive functions of their design.

User Interface Design

The effective of the UI design is also an important factor contributing to the effectiveness of the representation in terms of in wayfinding for users. The most important metric to evaluate UI design is usability, which measures how well users can use the functionality of the UI (6.813/6.831: User Interface Design & Implementation). The dimensions of usability include learnability, efficiency and safety.

Learnability evaluates if the UI is easy to learn. Users learn using different approaches. Some users learn by exploring the interface for features to achieve their goals, some users learn by searching for solutions to their problems in the documentation of the interface, and some users learn by watching other people use the interface. A learnable UI should have user-friendly interactive visual cues to aid people's recognition of knowledge through various learning approaches (6.813/6.831: User Interface Design & Implementation). Common learnable interaction styles used in UI design include *Menu and Forms* and *Direct Manipulation. Menu and Forms* interface presents a series of menus or forms to the user, and is used in many traditional websites (6.813/6.831: User Interface Design & Implementation). *Direct Manipulation* requires a continuous visual representation of the data objects of the interface where the user interacts using physical actions or labeled button presses and "rapid, incremental, reversible, immediately visible effects" is shown on the interface (6.813/6.831: User Interface Design & Implementation). Both *Menu and Forms* and *Direct Manipulation* put information in the world into the interface, making recognition easy for users. Other principles of learnability in

UI design include *consistency*, *affordances* and *feedback*. *Consistency* makes sure that similar things should look and act the same whereas different things should look different, allowing the user to "transfer their existing knowledge easily to a new UI" (6.813/6.831: User Interface Design & Implementation). *Affordances* refers to "the perceived and actual properties of a thing, primarily the properties that determine how the thing could be operated" (6.813/6.831: User Interface Design & Implementation). Since *affordances* are innately learned from experience, choosing properties suitable for specific experiences of the users in designing the UI makes the users learn faster. *Feedback* refers to how the system changes visibly when the user performs an action. The UI design of *feedback* enhances learnability when the user invokes a part of the interface, and the UI appears to respond in an expected way.

Efficiency evaluates that once learned, if the UI is fast to use. Common factors affect the efficiency of the UI. First is the way information is presented. Since on the cognitive level, elements of perception and memory are called chunks, and presenting the information in the format of chunks of familiar information makes it easier to perceive and remember for users. In addition, the sizes of the elements on the UI and the distances between the elements affect the efficiency of the UI to a great extent, since they affect the time that users move their mouses to perform multiple clicking and steering tasks (6.813/6.831: User Interface Design & Implementation). Another way to improve efficiency is the use of shortcuts, including adding keyboard shortcuts for commands, adding default inputs and pending delete for input boxes, offering recently-used or frequently-used choices, minimizing typing with autocomplete, using aggregation to allow the user to perform multiple tasks of the same type at the same time, and put all needed information and tools for a particular task within the user's easy reach (6.813/6.831: User Interface Design & Implementation).

Safety evaluates that if errors of the UI are few and recoverable. According to how they occur, errors can be classified into *slips and lapses* and *mistakes*, where the former is a failure to correctly execute a procedure, and the latter is a failure to use the correct procedure for the goal. Most cases of UI errors are *slips and lapses*, among which are three most common types of errors: *capture errors*, *description errors* and *mode errors* (6.813/6.831: User Interface Design & Implementation). A *capture error* occurs when "a person starts executing one sequence of actions, but then veers off into another (usually more familiar) sequence that happened to start the same way" (6.813/6.831: User Interface Design & Implementation); a *description error* occurs when "two actions are very similar. The user intends to do one action, but accidentally substitutes the other" (6.813/6.831: User Interface Design & Implementation); a *mode error* occurs when there are multiple modes of the UI and "the user tries to invoke an action that does not have the desired effect in the current mode" (6.813/6.831: User Interface Design & Implementation). Common techniques in UI design to prevent and recover errors are implementing clear, concise and polite confirmation dialogs and error messages, and allowing users to undo changes they recently made. Error prevention techniques specifically targeting *mode errors* include eliminating modes, increasing visibility of mode, using temporary modes and using disjoint action sets in different modes (6.813/6.831: User Interface Design & Implementation).

Exploring Cognitive Wayfinding Behavior

The first part of this thesis involves a subject experiment. The goal of the experiment is to evaluate currently available frameworks that use 3-dimensional media for key strengths and weaknesses of each representation in helping people navigate. The representations evaluated are 3D geographic models (henceforth, "3D") in the 3D mode of Google Maps, and street view panorama (henceforth, "Street View") in the street view mode of Google Maps. The experiment is divided into two rounds.

Preliminary Experiment

The first round of the experiment is a preliminary trial round to gain insights into the questionnaire design of the second round of the experiment, in terms of the questions' length, content, and level of complexity. In addition, the preliminary experiment serves to assess the participants' familiarity with the representations in Google Maps, especially the street view mode, since pre-experiment interviews show that most people do not use the street view version of Google Maps.

Methodology

The preliminary experiment involves 10 participants from the MIT student body, 5 of whom have an architecture or urban planning background. Participants are given 3 minutes of free exploration of the Sultanahmet Square (Appendix A) in Istanbul on Google Maps on their laptop browser, with an instruction to pay attention to the architectures and their relative positions. Afterwards, they are invited to complete a wayfinding quiz and a survey (Appendix

A). The quiz questions gauge the participants' knowledge of the relative positions between two architectures, sense of scale in virtual environment of Google Maps, and ability to identify and articulate a path between two landmarks. The metrics are designed to correspond to principles of directions, spatial orientation, and path integration, all of which are important cognitive processes of wayfinding discussed in Golledge (ch. 5-7). Finally, interviews with participants are also conducted to obtain feedback on the experiment, the wayfinding quiz and the Google Maps UI.

Results of the experiment are evaluated both quantitatively (reflected in the grades of way-identification quiz) and qualitatively (reflected in the feedback participants give in the interviews) to enlighten the design of the revised experiment.

Results

The way-identification quiz consists of three types of questions, grouped by abilities evaluated as direction questions, scale question and wayfinding question.

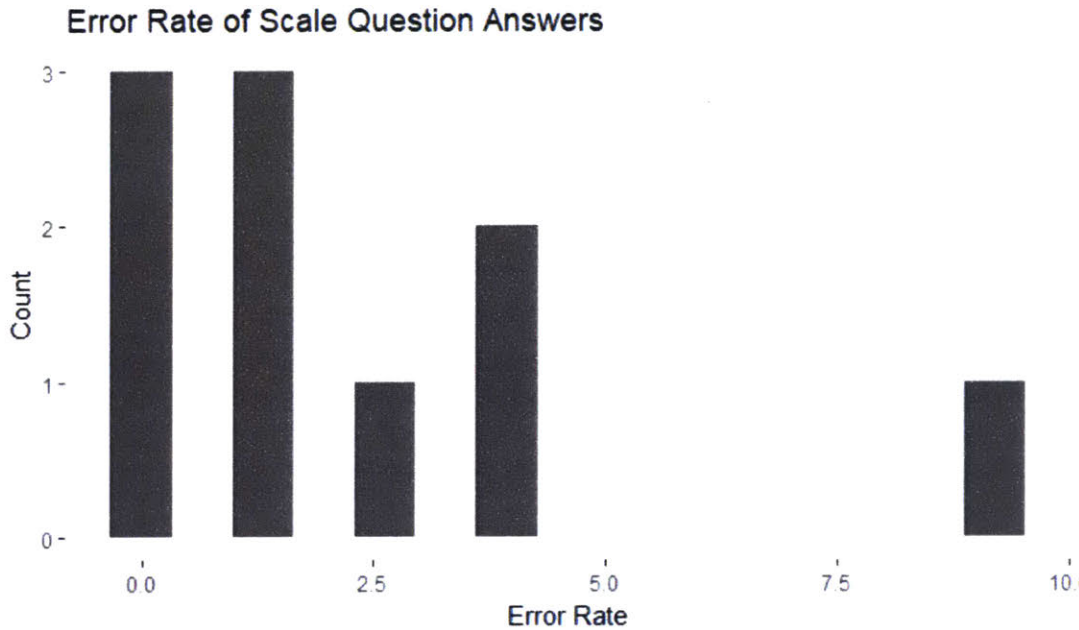
Table 1 shows the results of direction questions, which are three multiple-choice questions, each asking about the direction of one architecture with respect to another architecture. Each column in Table 1 displays the preliminary experiment statistics for each direction question, including the distribution of the participants' answers, and also the correctness both among the entire group of 10 participants and by gender and background (architecture/planning or non-architecture/planning).

		Question 1	Question 2	Question 3
Answer Distribution (correct answer highlighted)	A	1/10%	4/40%	4/40%
	B	1/10%	1/10%	2/20%
	C	7/70%	2/20%	0/0%
	D	1/10%	3/30%	4/40%
Answer Correctness	Entire Cohort	70%	30%	40%
	Male/Female	83.3%/66.7%	60%/0%	42.9%/33.3%
	AP/non-AP	60%/80%	20%/40%	60%/20%

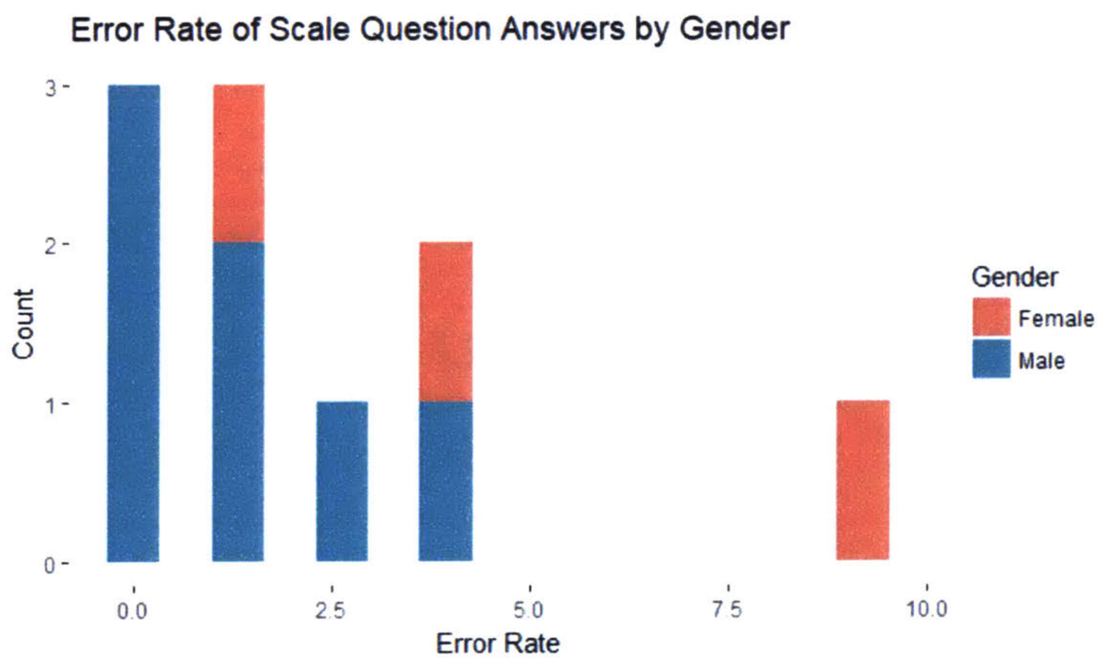
Table 1. Answer distributions and correctness (in percentage) of direction questions in the preliminary experiment.

Figure 3 shows the error rates for the answers for the scale question, which asks the participant to estimate the distance between two architectures on the map. The correct answer is within a range of 100-200 meters. The lower bound was calculated by measuring the direct distance between the two architectures using the scale on the map, while the upper bound was taken from the length of a path between the two architectures on Google Maps. The error rate was calculated for each of the participant's answer, and plotted in Figure 3 both as a distribution of the entire cohort (Figure 3a), and by gender (Figure 3b) and background (Figure 3c).

a


Error Rate of Scale Question Answers
3
2
1
0
Count
0.0
2.5
5.0
7.5
Error Rate

b


Error Rate of Scale Question Answers by Gender
3
2
1
0
Count
0.0
2.5
5.0
7.5
10.0
Error Rate
Gender
Female
Male

c

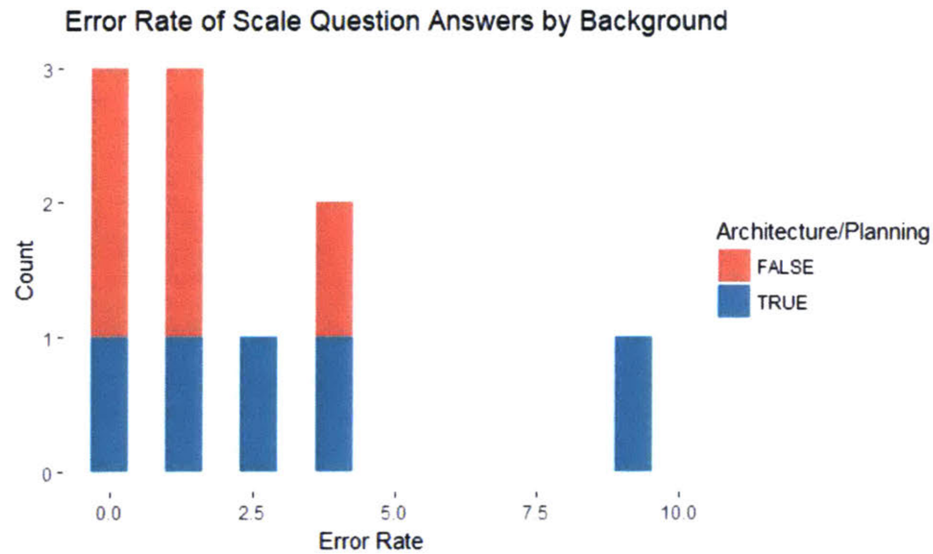
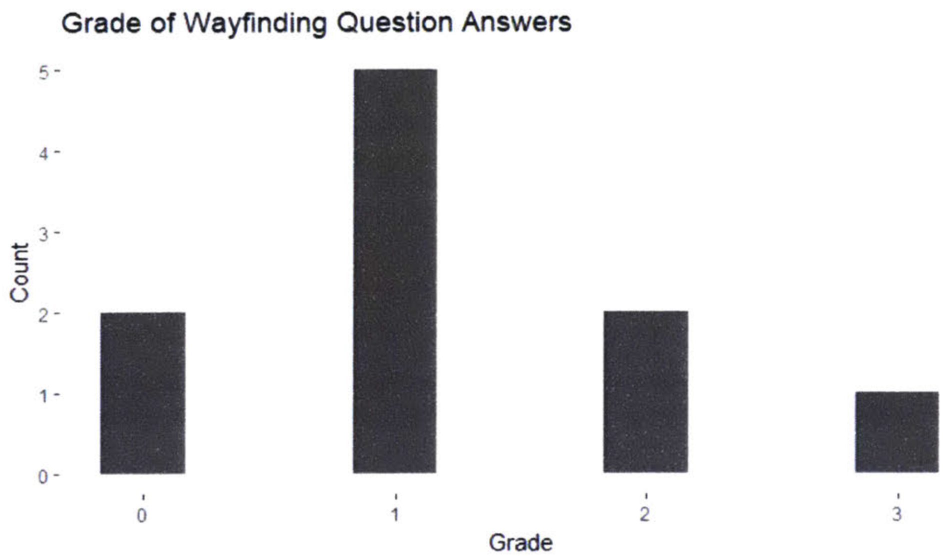


Figure 3. Participants' error rate distribution of scale question in the preliminary experiment, (a) among the entire cohort, (b) by gender, and (c) by background.

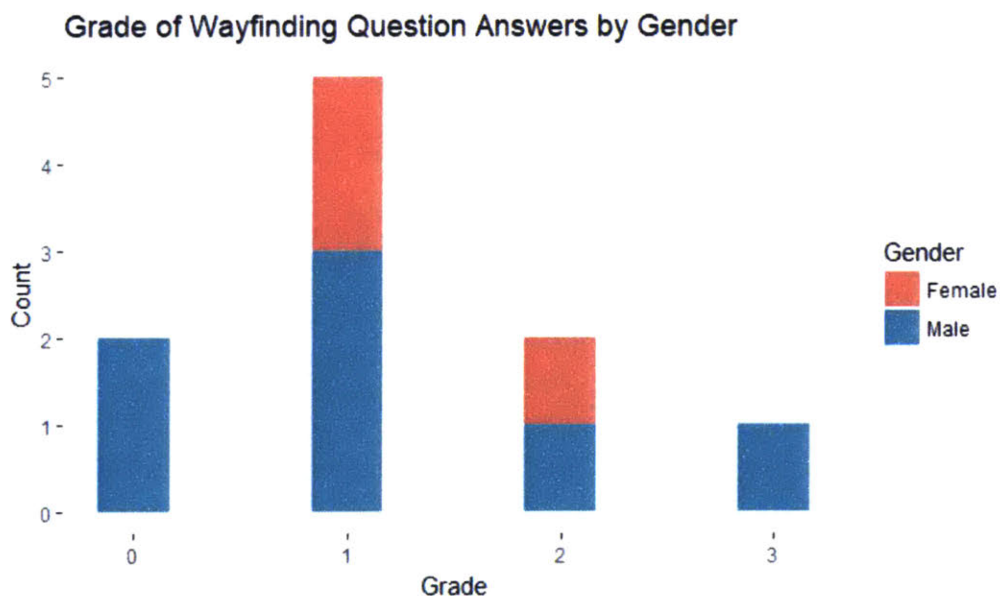
Figure 4 shows the grades for the answers for the wayfinding question, which asks the participant to describe a route between two architectures. The grading is based on a 0-3 point scale, where a 3-point answer completely, correctly and clearly describes all the information about the path between the two architectures, a 2-point answer correctly and clearly describes part of the information about the path between the two architectures, or completely and clearly describes all the information about the path with 1 or 2 small mistakes, a 1-point answer includes some correct information about the path between the architectures, and a 0-point answer does not include any correct information about the path between the two architectures. The grades are determined based on the rubric and plotted in Figure 4 both as a distribution of the entire cohort (Figure 4a), and by gender (Figure 4b) and background (Figure 4c).

a



Grade of Wayfinding Question Answers
Count
Grade
0
1
2
3
4
5

b



Grade of Wayfinding Question Answers by Gender
Count
Grade
Gender
Female
Male
0
1
2
3
4
5

c

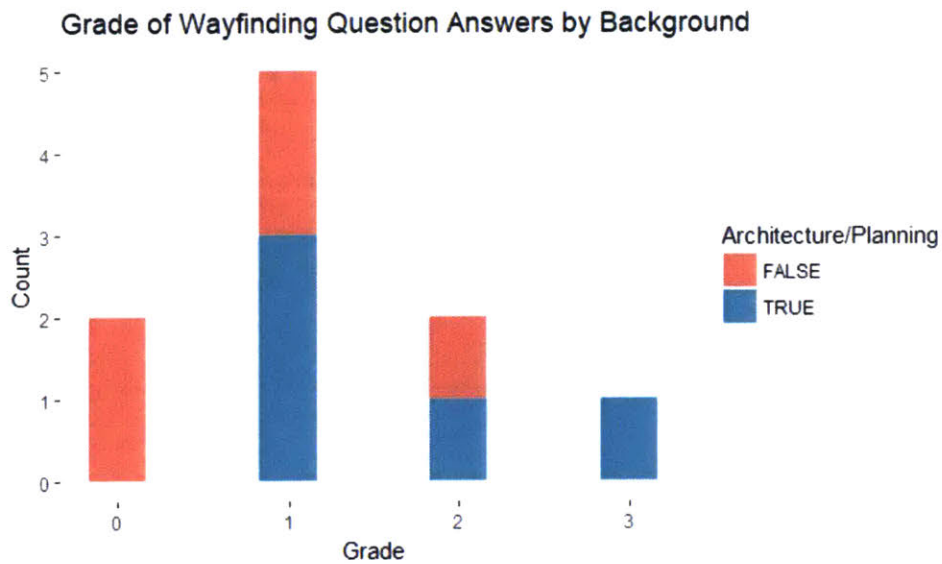


Figure 4. Participants' grade distribution of wayfinding question answers in the preliminary experiment, (a) among the entire cohort, (b) by gender, and (c) by background.

Discussion

According to Table 1, people tend to perform differently on identifying the relative directions between architectures. In addition, people perform differently on the three questions: whereas 70% of the participants answer Question 1 correctly, only 30% and 40% of the participants answer Question 2 and 3 correctly, respectively. The results are surprising since we expected a better overall performance and a smaller variation among the performances on the three questions. Another interesting discovery is that male participants tend to outperform female participants in all questions, and that there is no correlation between the participant's background and his/her performance on the three questions. During the interview with the participants, we found a few reasons that might have caused the unexpected results: a few people mentioned that they either did not remember where the architectures in Question 2 or Question 3 are, or found the style of architectures in the area too homogenous to tell apart from each other, and therefore could not map the architectures in Question 2 or Question 3 to their locations. However, those are not factors that reflect the lack of wayfinding knowledge and should not affect the results of the experiment. Therefore, the results show that the questions in the next round of experiment should avoid tasks that involves recognition and memorization of architectures so that they will not bias the results of the experiment.

Figure 3 shows that people tend to overestimate the distance between the two designated architectures on the map. Only 30% of the participants estimate the distance rather precisely (with an error rate within +/- 100% of the actual distance range). Similar to the previous type of questions, male participants tend to outperform female participants in learning the scale, with an overall lower error rate. Additionally, participants from an architecture/planning background do not seem to outperform their peers without an architecture/planning background. During the interview with the participants, we found that a number of participants use the street view mode of Google Maps and use the arrows to move around and explore the area within street view mode. However, because the arrows move more slowly than they expect, they all end up overestimating the speed the arrows move in real scene, thereby overestimating the distances on the map. This might explain the overall trend of the answer, though it does not seem to reflect a flaw in the design of the Google Maps UI that results in the lack of effectiveness of wayfinding. In fact, it is more likely related to the display format for the Google Maps UI -- virtual reality might be a better way for people to understand the movement in Google Maps since it is more immersive than the web app we are testing. While the display format is also a variable of influence that is worth studying, it is not the main focus of our research.

Figure 4 shows that people tend to have different performance on describing paths between architectures. Unlike the other two types of questions, there is not a clear distinction between the performance of male and female participants, and between the performance of architecture/planning and non-architecture/planning participants. Similar to the results of the direction questions, the results to the way-finding question are also worse than expected. From the interviews of the participants, we find that also similar to the direction questions, in tackling the way-finding questions, the participants require recognition and memorization of architectures that they may not do well. This biases their performance on their wayfinding task and should be eliminated from future experiments. In addition, another observation from the answers of wayfinding question is that people tend to describe paths in different ways. Whereas some people tend to describe paths using absolute directions (north, south, east and west), other people tend to describe paths with respect to their relative positions on the map. The former is known as *allocentric* spatial processing, which involves an object-to-object representational system, whereas the latter is known as *egocentric* spatial processing, which involves a self-to-object representational system. Different people have different abilities and preferences for their way of spatial processing, which is another variable worth noting in evaluating the effectiveness in the two different modes of representation in Google Maps. It seems that 3D mode might be more effective for people with good *allocentric* spatial abilities, since it clearly shows the directions on the map and the relative positions among the architectures. Street View mode, on the other hand, might be more effective for people with good *egocentric* spatial abilities, since it puts the participant in the scene so that he/she can learn the surrounding architectures better by experiencing their positions with his/her own position on the map.

Revised Experiment

The second round of the experiment is designed based on the quiz results and interview feedback of the preliminary experiment. Compared to the preliminary experiment, it is conducted on a larger scale and involves more test locations and types of evaluation questions.

Methodology

20 participants from the MIT student body are involved in the revised experiment. Similar to the preliminary experiment, half of the sample have an architecture or urban planning background, while the other half have other backgrounds. Participants are divided into two groups, each with people from both halves of the sample. A set of 4 locations are picked for the evaluation: Kyoto Station area in Kyoto, Japan; Vancouver City Centre area in Vancouver, Canada; Central area in Hong Kong, China; and Historic City Center area in Mexico City, Mexico (Appendix B). The locations are picked such that each has both 3D models and comprehensive street view panoramas available. The groups are assigned tasks with locations and representations as shown in Table 2. The evaluation consists of a way-identification quiz and a virtual scavenger hunt, as well as a feedback survey.

	Group 1	Group 2
Kyoto	3D	Street View
Vancouver	Street View	3D
Hong Kong	3D	Street View
Mexico City	Street View	3D

Table 2. Representation mode assignment for two experiment groups by city, in both way-identification quiz and virtual scavenger hunt tasks in the revised experiment.

Way-identification Quiz

Each participant is allowed 3 minutes in each location, and is allowed only in his/her designated mode, not any other mode of Google Maps. Afterwards, a quiz is assigned to the participant, and the quiz consists of two multiple-choice questions that focus on assessing the participant's path integration and landmark-place association (Golledge, ch. 8).

Virtual Scavenger Hunt

Each participant is given an image of a specific target architecture within the location. Afterwards, each participant is allowed 5 minutes in each location, and is allowed only in his/her designated mode, not any other mode of Google Maps, attempting to find the target architecture. The participant's activity in his/her designated mode is timed and screen-recorded.

Feedback Survey

The feedback survey is given to each participant after a sequence of his/her way-identification and virtual scavenger hunt tasks. It asks the participant's feedback on the user experience in the framework, as well as prior experience in architecture, wayfinding and prior knowledge of Google Maps and the locations tested.

The results of the experiment are evaluated both quantitatively (reflected in the grades of way-identification quiz and time and accuracy in the virtual scavenger hunt) and qualitatively (reflected in the feedback participants give in the surveys).

Results

Way-identification Quiz

Scores for way-identification quizzes are calculated for all participants on all locations they were tested on. Based on their correctness on the two multiple choice questions, each of the three parts of Question 1 are worth 2 points, and people can get 2 points for completely correct answers, 1 point for partially correct answers, and 0 point for incorrect answers; Question 2 is worth 3 points, 1 point for each of the choice correctly selected/unselected. Since each participant was tested on two locations in 3D and the other two locations in Street View, the two scores in the same mode are added together for both Question 1 and Question 2 scores. After discarding scores of 0, we calculate the means and standard deviations (SDs) for the two questions in the two frameworks, among the entire group of 20 participants and by gender and background (architecture/planning or non-architecture/planning). The means and SDs of the scores are shown in Table 3.

		Question 1 Score in Street View		Question 1 Score in 3D		Question 2 Score in Street View		Question 2 Score in 3D	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Entire Cohort		7.86	1.75	9.20	1.52	3.86	1.03	3.93	1.22
by Group	Group 1	6.88	1.46	9.38	1.30	3.88	0.99	3.75	1.16
	Group 2	9.17	1.17	9	1.83	3.83	1.17	4.14	1.35
by Gender	Male	8	1.34	9.42	1.62	3.73	1.01	4.17	1.19
	Female	7.33	3.21	8.33	0.58	4.33	1.15	3.00	1.00
by Background	Architecture/ Planning	8	1.87	8.67	0.82	3.4	0.89	3.33	0.82
	Non-Architecture/Planning	7.78	1.79	9.56	1.81	4.11	1.05	4.33	1.32
by Spatial Processing Type	Egocentric	8.29	1.80	9.00	1.51	3.57	1.13	3.88	1.25
	Allocentric	7.43	1.72	9.43	1.62	4.14	0.90	4.00	1.29

Table 3. Means and SDs of participants' grades in 3D and in Street View on each of the two way-identification quiz questions of the revised experiment. The means and SDs are calculated among the entire cohort, by gender, by background, and by spatial processing type.

Virtual Scavenger Hunt

Similar to the results of the way-identification quiz, each participant takes part in the scavenger hunts at two locations in 3D and the other two locations in Street View. During each scavenger hunt activity, we record the time it took for each participant to find the target at all of the locations, and add together in the same mode. We also calculate both the means and standard deviations (SDs) for the times in the two frameworks, among the entire group of 20 participants and by gender and background (architecture/planning or non-architecture/planning). The means and SDs of the scores are shown in Table 4.

		Wayfinding Time in Street View (s)		Wayfinding Time in 3D (s)	
		Mean	SD	Mean	SD
Entire Cohort		147.00	125.73	164.67	103.70
by Group	Group 1	138.13	67.36	117.50	71.38
	Group 2	158.53	185.62	218.57	113.11
by Gender	Male	159.09	140.29	175.58	110.37
	Female	102.67	24.44	121	68.61
by Background	Architecture/Planning	147.4	57.89	155.17	83.64
	Non-Architecture/Planning	146.78	154.96	171	119.73
by Spatial Processing Type	Egocentric	139.86	69.43	176.00	115.95
	Allocentric	154.14	171.21	151.71	95.09

Table 4. Means and SDs of participants' recorded time in 3D and in Street View to find the target landmark in the virtual scavenger hunt of the revised experiment. The means and SDs are calculated among the entire cohort, by gender, by background, and by spatial processing type.

Feedback Survey

After reading the responses from the participants on the feedback survey, we summarize a few arguments from the participants below:

1. 4 participants report that the 3D mode does not provide much more information than the 2D satellite map, whereas 2 participants report that the 3D mode provides important additional information about the terrain and very tall landmarks (e.g., skyscrapers), which are both very effective in assisting wayfinding.

2. 6 participants report that the 3D perspective is very different from eye-level perspective, and the 3D geographical models are in very low resolution, both making it very difficult to identify landmarks in 3D.
3. 10 participants report that the 3D mode does a good job in helping them get an overview of the place, making it easy for them to identify relative directions between landmarks.
4. 2 participants report that while wayfinding in 3D is easy in grid-based cities, it gets significantly more difficult when city structures get more complicated and irregular.
5. 16 participants report that the Street View mode does not give them a good sense of which direction they are walking to.
6. 4 participants report that the Street View mode is helpful for wayfinding. Elements like text and direction of cars that are either unclear or non-existent in 3D help them identify landmarks and directions.
7. 3 participants report that for locations where the buildings are very dense, it is more difficult for them to recognize landmarks in Street View.
8. 2 participants report that while it is easier in Street View for people to identify accessible paths (roads, footbridges, hallways etc.) and plan travels, it is not as easy as 3D to locate landmarks by their relative directions.
9. 1 participant reports that he encodes information differently in 3D and in Street View. While in both modes, he starts with identifying landmarks, in

3D he tends to remember direction and number of blocks between landmarks, whereas in Street View he tends to remember paths using other landmarks (e.g., turn at a specific landmark).

Discussion

Our results show a marked contrast among the characteristics studied. Gender makes no significant difference in wayfinding performances. Perhaps more surprisingly, participants with backgrounds in architecture and planning also show no distinctive performances. However, other factors were highly significant, as we now outline.

City Structure and Preference of Mode in Wayfinding

Table 3 shows that on average, Group 2 members outperform Group 1 in answering question 1 in Street View and answering question 2 in 3D, and Group 1 members outperform Group 2 in answering question 1 in 3D and answering question 2 in Street View. Since Group 1 explored Kyoto and Hong Kong in 3D, and Vancouver and Mexico City in Street View, whereas Group 2 explored explored Kyoto and Hong Kong in Street View, and Vancouver and Mexico City in 3D, the results show that for Kyoto and Hong Kong, landmark-place association is easier in 3D, whereas path integration is easier in Street View; for Vancouver and Mexico City, landmark-place association is easier in Street View, whereas path integration is easier in 3D.

The quantitative results align with the results from the feedback survey: both Kyoto and Hong Kong have a variety of different building heights in the target area, which Vancouver and Mexico City do not have. Therefore, the 3D mode makes it easy for participants to identify landmarks when their heights stand out in the civic backdrop. In addition, since Vancouver and Mexico City are both grid-based, it is easier to do path-integration in 3D. On the other hand, Hong Kong has an irregular city structure, which is not grid-based. In addition, it has footbridges that are much easier to identify in Street View. Therefore, it is easier to do path-integration in Street View.

Spatial Processing Type and Preference of Mode in Wayfinding

Both Table 3 and Table 4 show that, on average, people who are more used to egocentric spatial processing have higher performance in Street View in all tasks with the exception of answering question 2, whereas people who are more used to allocentric spatial processing have higher performance in 3D in all tasks. These results suggest that the 3D mode is better for allocentric spatial processing, whereas the Street View mode is better for egocentric spatial processing.

The quantitative results also align with the results from the feedback survey, as well as with our hypothesis. The 3D mode gives the users a better overview of the city, and therefore gives them a better sense of direction between landmarks; the Street View mode helps the users identify paths between landmarks, and therefore gives them a better sense of how they can get from one landmark to the other. The former involves object-to-object representation, which is more relevant to allocentric spatial processing, whereas the latter involves a self-to-object representation, which is more relevant to egocentric spatial processing.

Task Type and Preference of Mode in Wayfinding

Table 3 shows that, on average, the 3D mode outperforms the Street View mode in helping people answer both questions on the way-identification quizzes, which aligns with the participants' conclusion that the 3D mode does a good job in helping them get an overview of the place, making it easy for them to identify relative directions between landmarks. The results suggest that getting an overview of the place on a map is integral to the cognitive behaviors of wayfinding, including path integration and landmark-space association.

Table 4 shows that, on average, the Street View mode outperforms the 3D mode in helping people perform the virtual scavenger hunt tasks. This result aligns with the participants' conclusions, that compared to the low-resolution models in the bird's eye view in the 3D mode, the Street View mode is more similar to the real feeling of walking along a street. Therefore, it is easier for the participants to identify target landmarks in Street View than in 3D. However, the SDs in Table 4 show that there is a higher variance in performance among the participants in Street View, compared to that in 3D. This result also aligns with the participants' conclusion that Street View makes it hard for them to know geographic directions (north, south, etc.). Therefore, participants might easily get lost in Street View, but not in 3D. In addition, since 3D gives the participants an overview of the location, which Street View does not, the time it takes for participants to find the target landmark depends greatly on the direction they are walking to in Street View. Without prior knowledge of the cities, the participants choose the directions randomly, leading to a higher variation in their performances on the virtual scavenger hunt task.

Creating a Framework for User Interface Design

Following the subject experiment is the development of a framework, tested with a prototype for a UI that aims to improve the navigation experience for users.

Methodology

Based on the recurring themes in the results from the subject experiment, we find that the two representations have different strengths in helping people perform wayfinding tasks. The major strength of the 3D mode is that it gives the users an overview of the place, making it easy to identify directions; the major strength of the Street View mode is that it gives the users a realistic image of the surroundings similar to that when the users are walking down the street. Incorporating the key strengths of the two frameworks, we propose a UI that has street view facades added in order to augment the 2D satellite map.

We create the framework by using Google Street View Image API to scrape the street view images at orthogonal angles to the streets at all the locations on the street where street view panoramas were taken ("Street View Image API"). To obtain the facades of buildings of different heights, at each street view panorama location, we scrape the street view images at multiple pitches and stitch the images together. After obtaining the facade images at all the street view panorama locations, we stitch them together to make street view facades, which we add to the 2D satellite map. For this thesis, we develop a proof-of-concept prototype of the UI of Cinco de Mayo, the street in Mexico City that we used as one of the locations in the revised experiment.

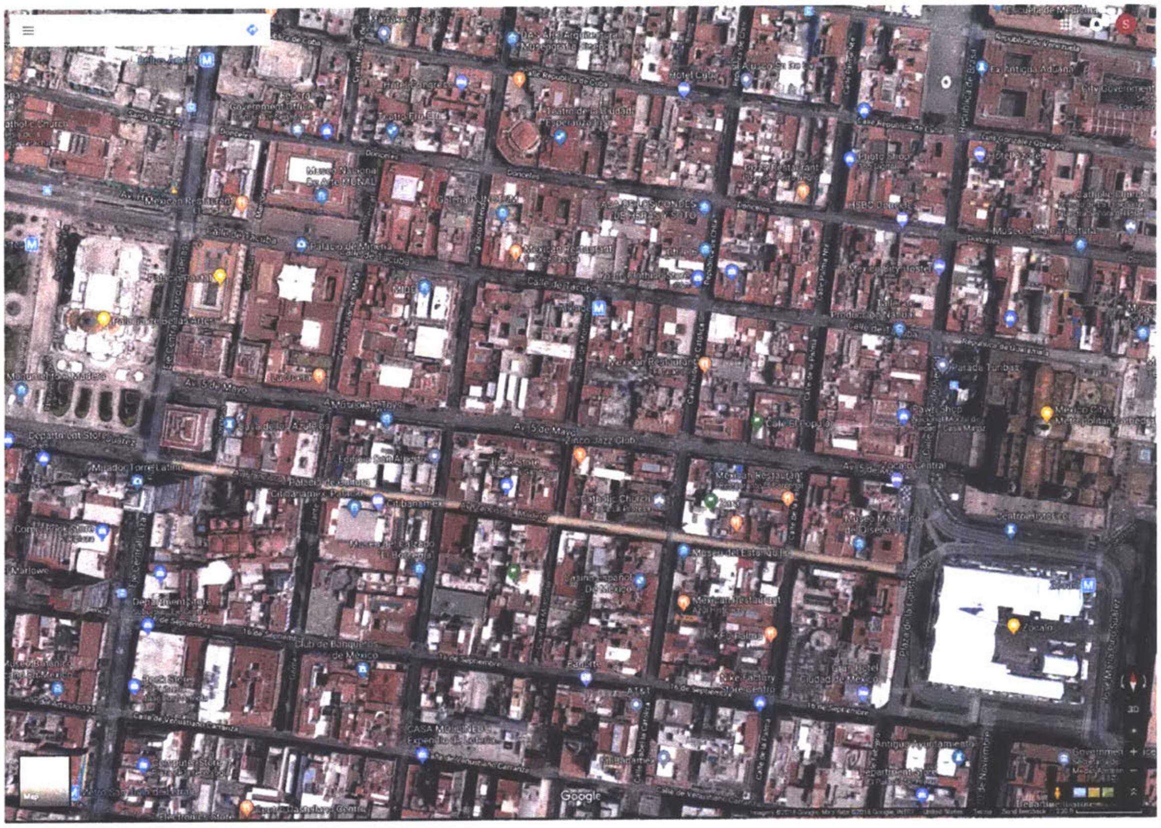
Analysis

Figure 5 shows a few screenshots of the UI prototype with street view facades augmenting the 2D satellite map of Cinco de Mayo (Figure 5a). A few design decisions are made in developing this prototype: for better learnability, when the mouse hovers over the group of buildings in the facade image, the outline of those buildings become highlighted (Figure 5b). After clicking, the facade image is displayed in place of the group of buildings (Figure 5c). For efficiency, clicking on the street displays the facade images on both sides of the section of the street closest to the click position (Figure 5d). As with the current version of Google Maps, the user can freely rotate and scale the view, and the facade images also rotate or scale with the view (Figure 5e). Clicking on the facade images hides the images in the 2D satellite view.

The goal of the UI is to combine the advantage of the 3D mode, which gives users an overview of the city, and the advantage of the Street View mode, which gives users a realistic image of the landmarks in the city. To reinforce the correlations between the landmarks in plan satellite view and in elevation facade view, we use the principle of proximity in UI design and decide to display the facade images in place of the buildings in 2D satellite view, so that users can cognitively map the two types of representation together in learning and memorizing specific landmarks. The design of the UI also takes into account the principle of universality in UI design. It accommodates the needs of both egocentric and allocentric users. Egocentric users are able to rotate the map and see rotated facades based on the view, allowing them to recognize nearby landmarks with respect to their own specific locations on the map. Allocentric users are able to identify relative directions between landmarks on the 2D map, and check the facade image to get a better sense of what the landmarks look like.

We also recognize a few limitations of the prototype. The resolution of the images obtained from Google Street View APIs is not as high as that of the original street view panoramas. The alignment of the images is also imperfect due to errors from panorama-taking and from stitching. Both the low resolution and the flawed alignment of the images result in low quality of the facade images augmenting the 2D satellite map. In addition, the current implementation only works for cities with straight roads, but not cities with complicated and irregular road structures such as Hong Kong. While the former problem can be solved with support from Google and computer vision software that performs auto-correction for image stitching, solving the latter problem requires more thorough research and design iterations to create a more versatile UI framework.

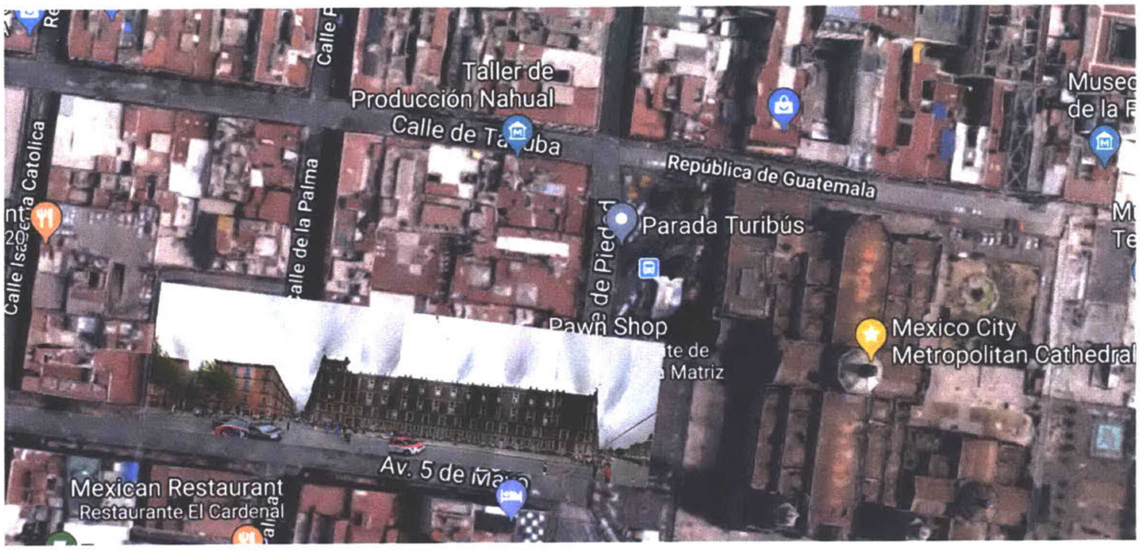
a



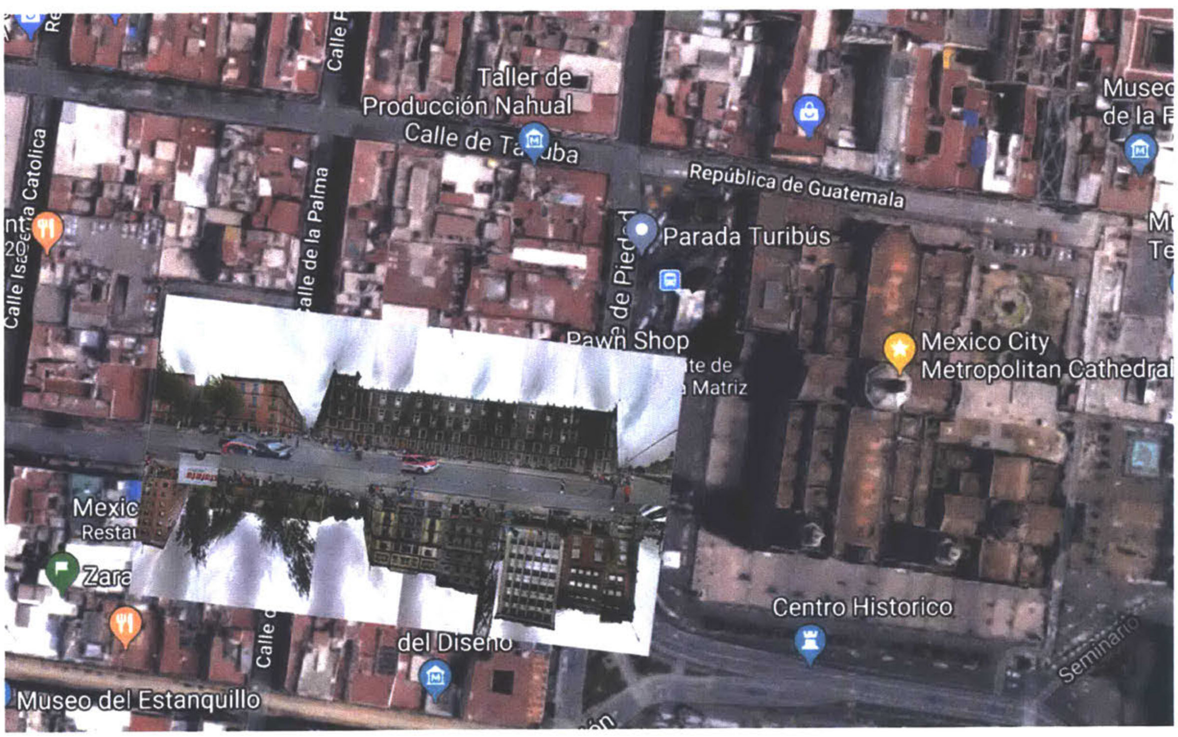
b

Taller de
Producción Nahual
República de Guatemala
Parada Turibús
Calle de la Palma
Pawn Shop
Nacional Monte de
Piedad - Casa Matriz
Café El Popular
Mexico City
Metropolitan Cathedral
Zócalo Central
Mexican Restaurant
Restaurante El Cardenal

c


Taller de
Producción Nahual
República de Guatemala
Parada Turibús
Pawn Shop
Mexico City
Metropolitan Cathedral
Mexican Restaurant
Restaurante El Cardenal

d


Taller de
Producción Nahual
República de Guatemala
Parada Turibús
Pawn Shop
Mexico City
Metropolitan Cathedral
Centro Historico
del Diseño
Museo del Estanquillo
Seminario

e

f

Figure 5. Screenshots of the proposed UI framework for wayfinding at the location of Cinco de Mayo, Mexico City. (a) shows the original 2D satellite map, from "Mexico City, Mexico." Map, *Google Maps*. Accessed 23 Apr. 2018; (b) shows highlighted building contours on hovering on the buildings; (c) shows facade image of the buildings on clicking on the buildings; (d) shows facade images of both sides of the street on clicking on the street; (e) shows a rotated and scaled view of the 2D satellite map with accordingly rotated and scaled facade images; (f) shows an enlarged view of the facade images augmenting the 2D satellite map.

Conclusion

This thesis studies factors influencing people's cognitive wayfinding behavior by using a comparative approach. By comparing the 3D mode and the Street View mode of Google Maps, we find that city structure, spatial processing type preference, and wayfinding task type affect people's preferences of mode in wayfinding. (Gender and a background in architecture or planning do not make a significant difference.) We also find that the major advantage of the 3D mode is the overview it gives, helping users orient themselves, and the major advantage of the Street View mode is the realistic image it shows, which helps users recognize landmarks. Based on the results from the subject experiment, we design a UI that has street view facades augmenting the 2D satellite map. Further research can be done to scale up the subject experiment, such as conducting the experiment not only with MIT students, but also with a more diverse sample of participants. In addition, other variables could also be involved in the experiment, such as age and cultural background, and they might also be significant factors in assisting wayfinding. In the future, the UI design and development methodology in this thesis can be refined to adjust to a greater variation of city structure types.

References

Bliss, James P., Tidwell, Philip D., and Guest, Michael A. "The effectiveness of virtual reality for administering spatial navigation training to firefighters." *Presence: Teleoperators and Virtual Environments* 6.1 (1997): 73-86.

"Cambridge USA." *Visit Cambridge Massachusetts | Cambridge Office of Tourism*, www.cambridgeusa.org/.

Choi, Joshua. *Merging Three Spaces: Exploring User Interface Framework for Spatial Design in Virtual Reality*. Diss. Massachusetts Institute of Technology, 2016.

Golledge, Reginald G., ed. *Wayfinding behavior: Cognitive mapping and other spatial processes*. JHU press, 1999.

Google Maps, Google, maps.google.com/.

Heydarian, Arsalan, et al. "Immersive virtual environments versus physical built environments: A benchmarking study for building design and user-built environment explorations." *Automation in Construction* 54 (2015): 116-126.

Lynch, Kevin. *The image of the city*. Vol. 11. MIT press, 1960.

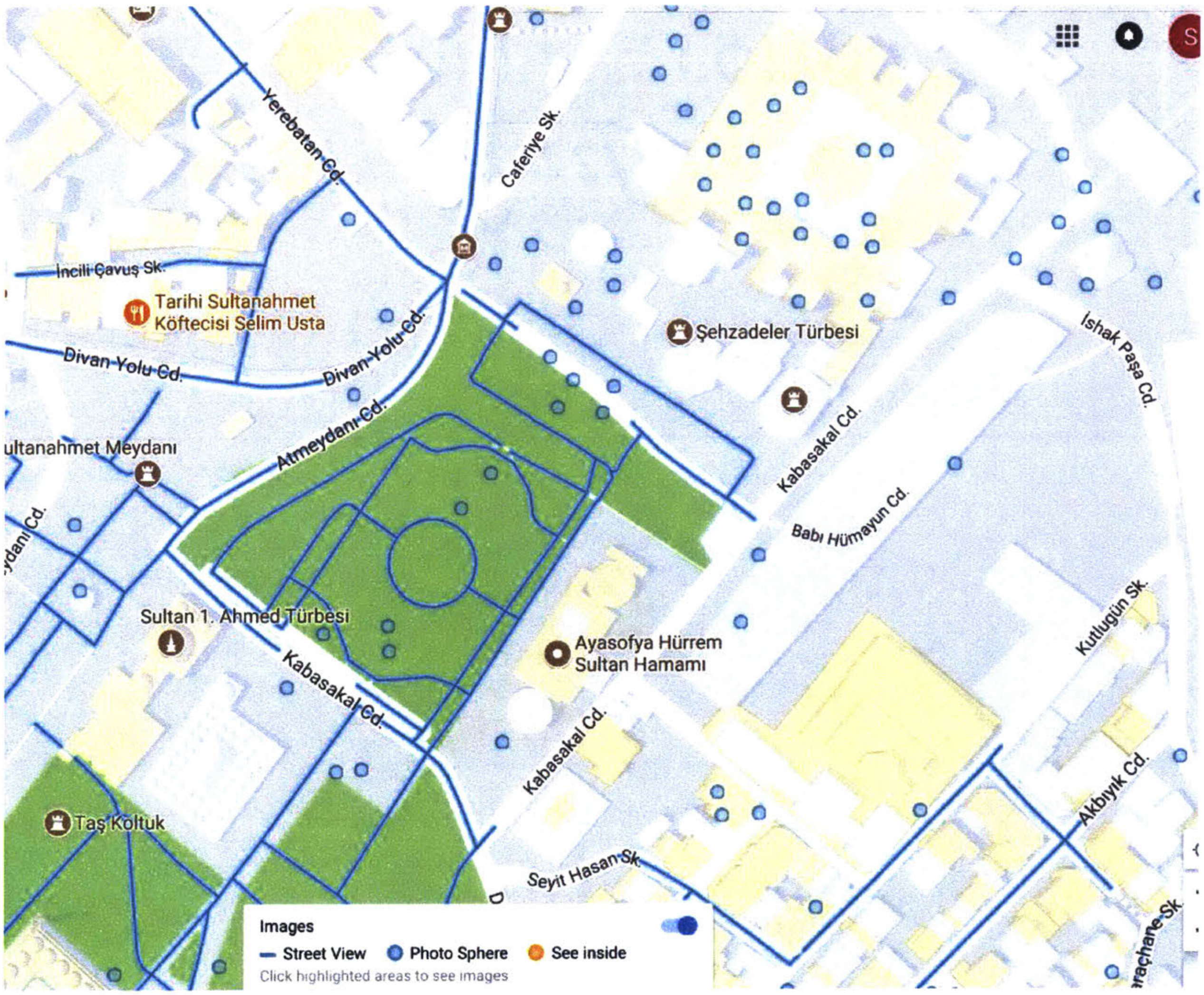
"Street View API." *Google Maps Platform*, Google, developers.google.com/maps/documentation/streetview/intro.

6.813/6.831: User Interface Design & Implementation, web.mit.edu/6.813/www/sp18/.

Appendices

A. Task Briefings of Wayfinding Quiz in the Preliminary Experiment

Each participant is allowed 3min in Google Maps. Beforehand, the participants are notified that they should explore in the area of interest and will be asked to complete a wayfinding quiz about the area. However, they are not asked to look for specific things in the area. Note-taking is not allowed in the process.



Wayfinding Quiz

For your reference, each architecture referenced in the questions is labelled below, with its name and facade photo.

Label	Photo	Name
a		Haseki Hurrem Sultan Hamam
b		Sultan Ahmad Maydan Fountain
c		Istanbul's Sultanahmet Foundation
d		Hagia Sophia

Direction Question

What is the rough direction of a, c and d with respect to b? Choose from the following:

A. Northeast B. Northwest C. Southeast D. Southwest

Direction of a with respect to b	
Direction of c with respect to b	
Direction of d with respect to b	

Scale Question
Estimate how long the path is from a to d (in m or ft, please write unit of your choice): ______

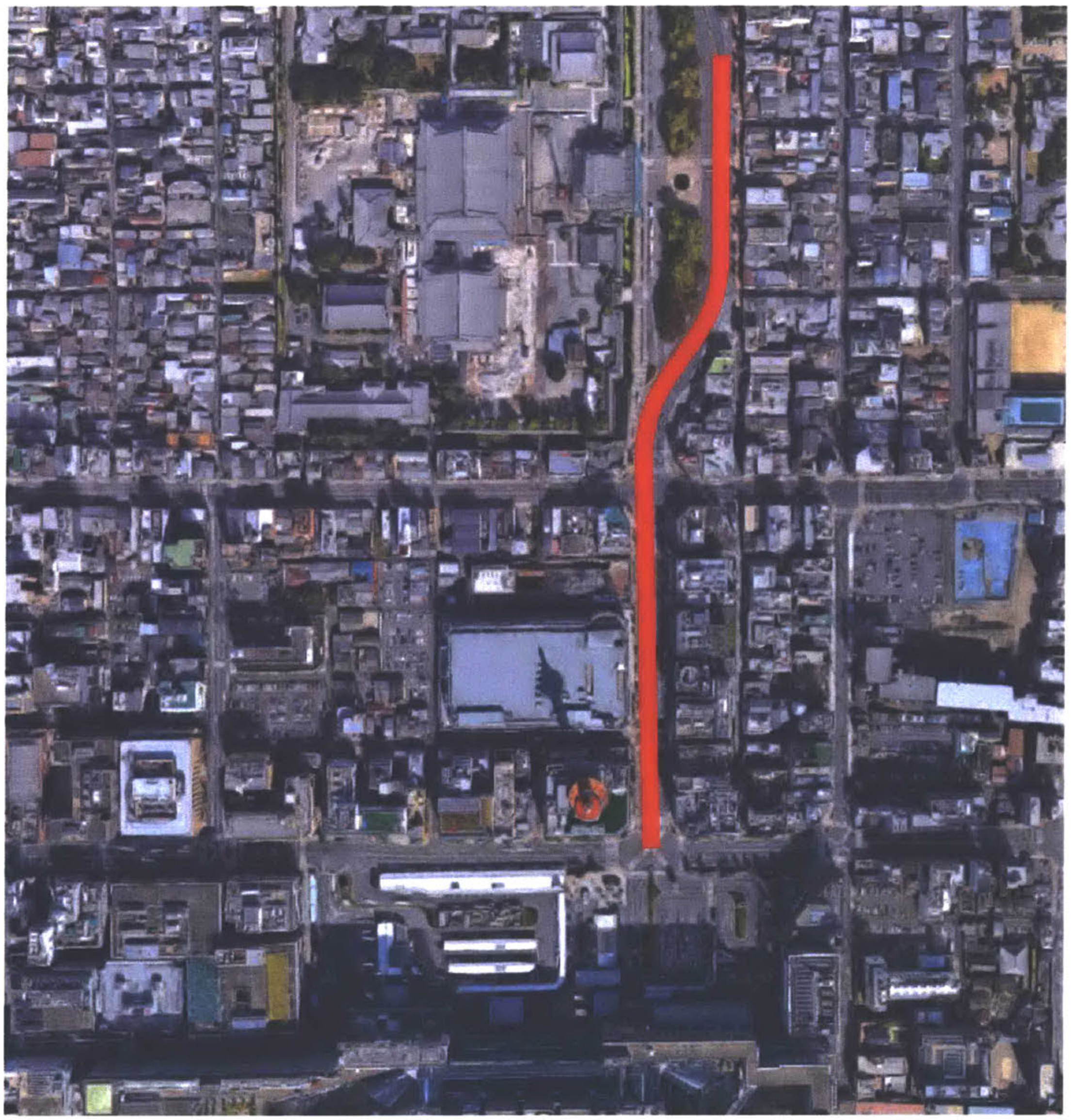
Wayfinding Question

Describe one path from c to d with as many details as possible. You can use directions (north/south/east/west), left/right turns, description of possible landmarks to describe the path.

B. Task Briefings of Way-identification Quiz and Virtual Scavenger Hunt in the Revised Experiment

Kyoto

3D:
https://www.google.com/maps/@34.9886966,135.759287,747m/data=!3m1!1e3
Street View:
https://www.google.com/maps/@34.9872459,135.7595946,3a,75y,17.95h,105.69t/data=!3m6!1e1!3m4!1sXZiqvosfvVWP2ko_NK4L_w!2e0!7i13312!8i6656



Way-identification Quiz:

Label	Photo	Name
a		Kyoto Tower
b		Higashi Honganji Temple
c		Kyoto Station

Q1. Look at the picture. If you are the person taking the picture, a/b/c is on your ______.
(Choose either one answer or two answers, one A/B and one C/D)

A. Left
B. Right
C. Front
D. Back

a	
b	
c	

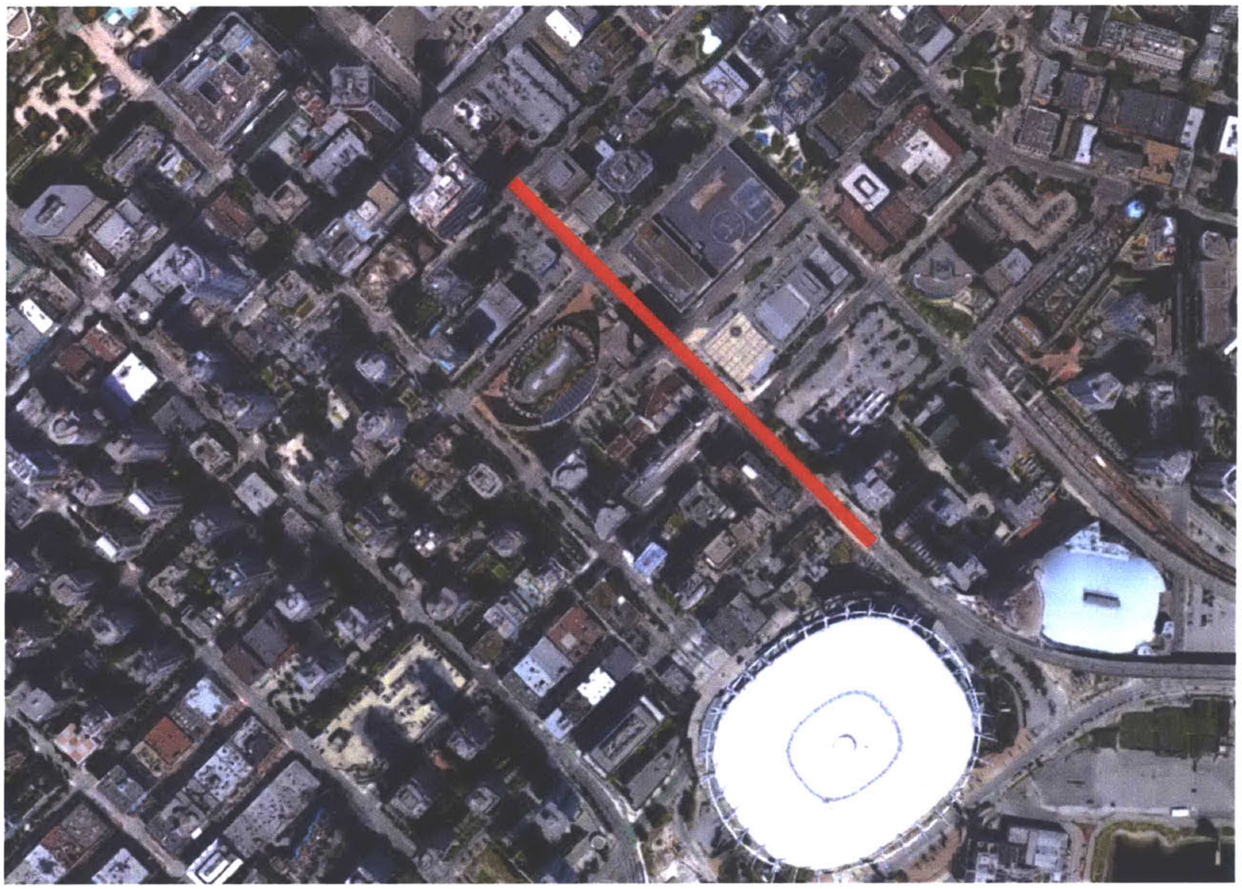
Q2. Choose all correct answers.

A. c is ~5min walk south of a.
B. b is ~10min walk south of c.
C. b is ~700m north of a.

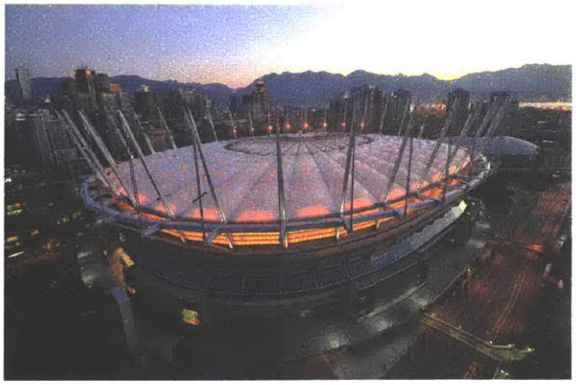
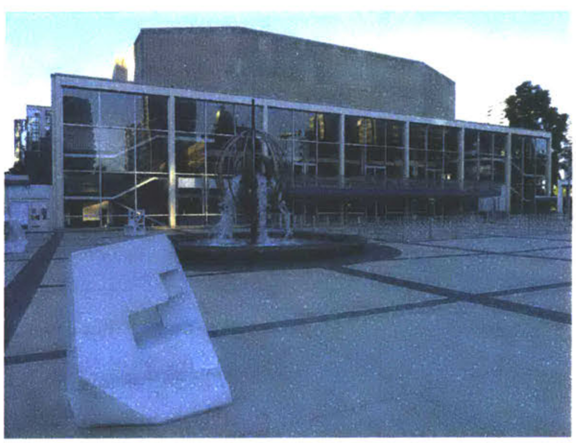
Virtual Scavenger Hunt (find this architecture):

Vancouver

3D:
https://www.google.com/maps/@49.2792208,-123.1132709,936m/data=!3m1!1e3
Street View:
https://www.google.com/maps/@49.2786658,-123.112382,3a,75y,346.69h,80t/data=!3m6!1e1!3m4!1s_YRFiktkI51LO8UaBCYerA!2e0!7i13312!8i6656



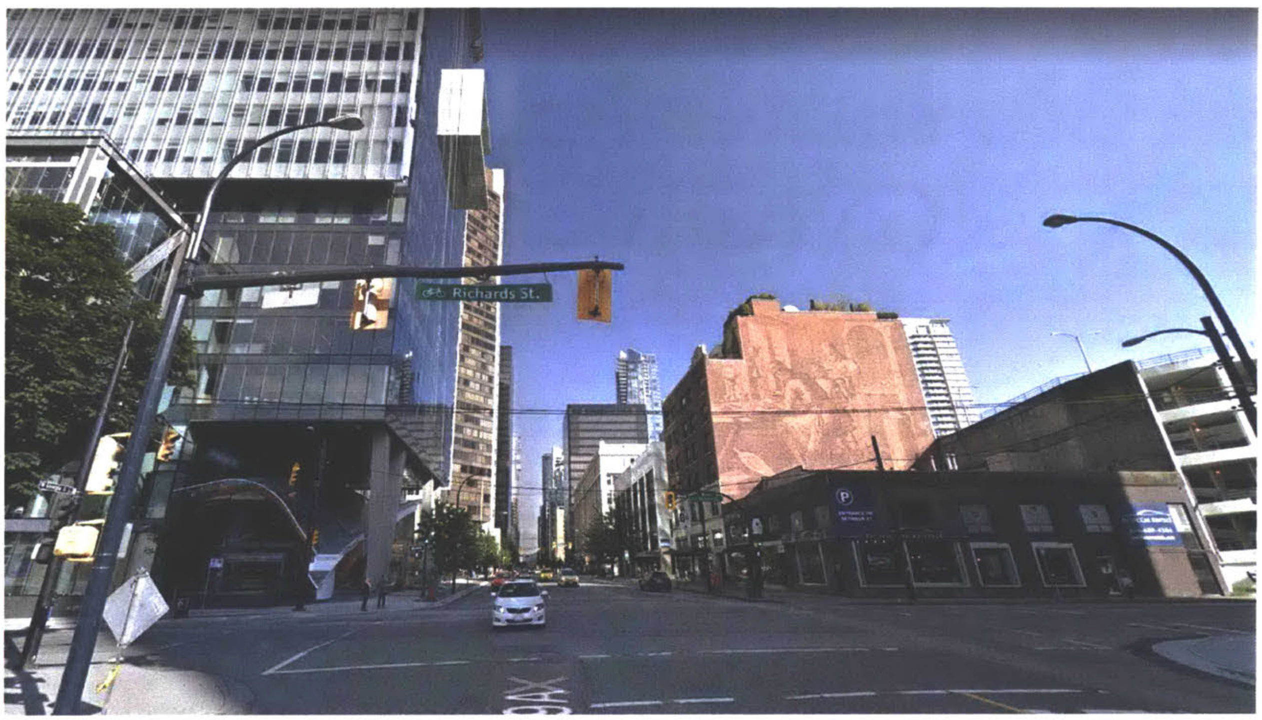
Way-identification Quiz:

Label	Photo	Name
a		BC Place
b		Vancouver Public Library
c		Queen Elizabeth Theatre

Q1. Look at the picture. If you are the person taking the picture, a/b/c is on your ______.
(Choose either one answer or two answers, one A/B and one C/D)

A. Left
B. Right
C. Front
D. Back

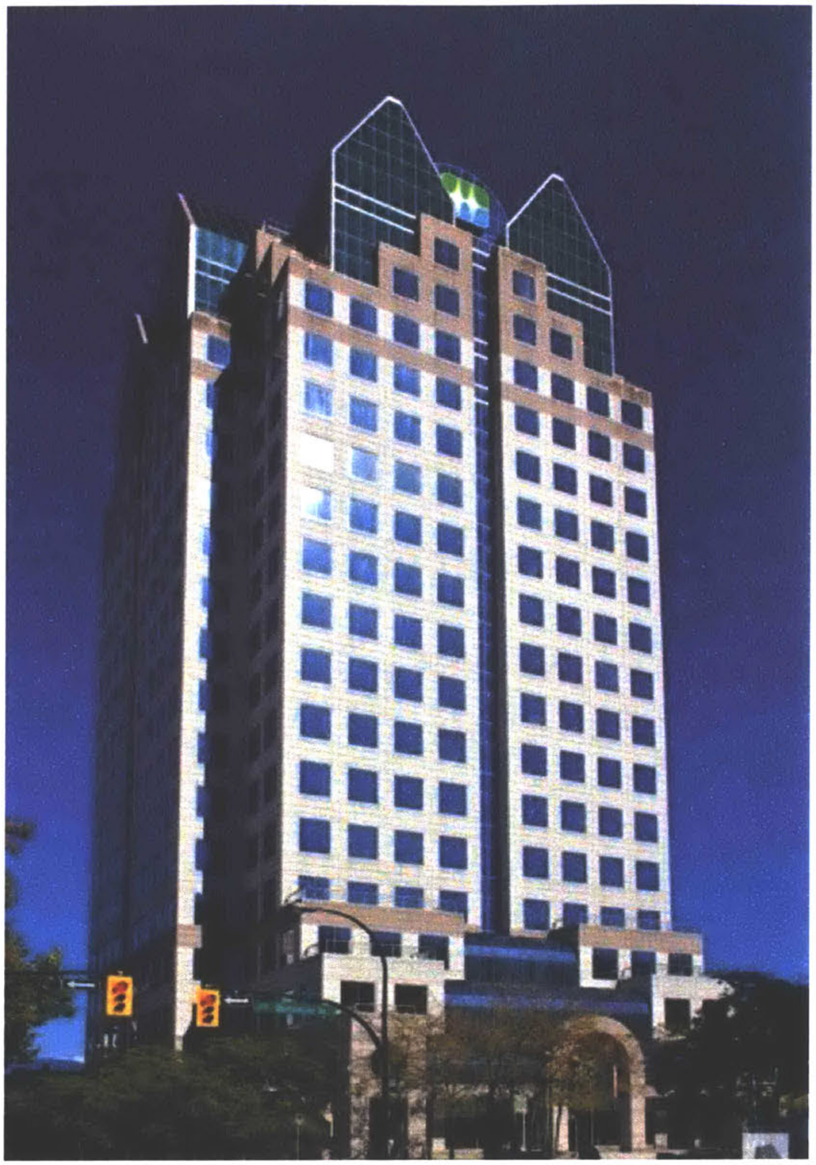
a	
b	
c	



Q2. Choose all correct answers.

A. c is ~5min walk east of b.
B. a is ~500m southeast of b.
C. a is ~500m southeast of c.

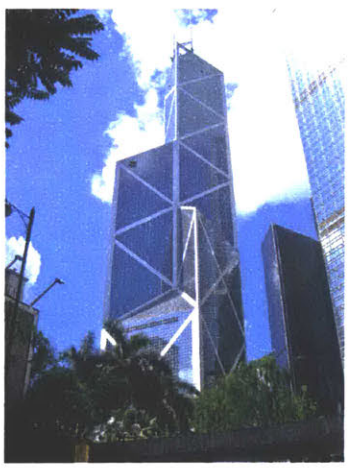
Virtual Scavenger Hunt (find this architecture):



Hong Kong

3D:
https://www.google.com/maps/@22.2792731,114.1632383,826m/data=!3m1!1e3
Street View:
https://www.google.com/maps/@22.2779087,114.1670322,3a,75y,282.03h,86.37t/data=!3m6!1e1!3m4!1sTeiDVc7y_Emk9cyjpPF9_g!2e0!7i13312!8i6656

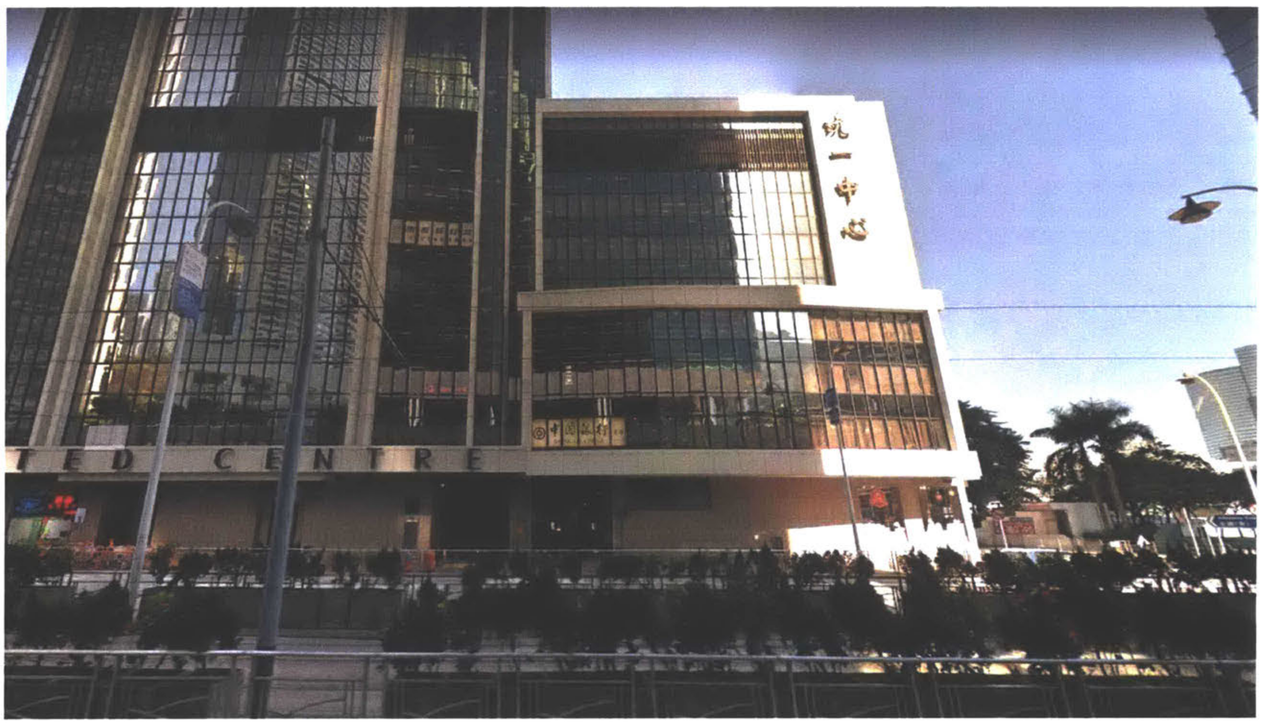
Way-identification Quiz:

Label	Photo	Name
a		Lippo Centre Towers
b		Bank of China Tower
c		CITIC Tower

Q1. Look at the picture. If you are the person taking the picture, a/b/c is on your ______.
(Choose either one answer or two answers, one A/B and one C/D)

A. Left
B. Right
C. Front
D. Back

a	
b	
c	



Q2. Choose all correct answers.

A. c is ~1000m north of b.
B. b is ~7min walk west of a.
C. To walk from c to a, one can walk ~300m north and then ~400m west.

Virtual Scavenger Hunt (find this architecture):

Mexico City

3D:
https://www.google.com/maps/@19.4345199,-99.1370701,759m/data=!3m1!1e3
Street View:
https://www.google.com/maps/@19.4339793,-99.1341144,3a,75y,283.59h,81.43t/data=!3m6!1e1!3m4!1snN2r4wF1FNT6d6-MMjVJsQ!2e0!7i13312!8i6656

Way-identification Quiz:

Label	Photo	Name
a		Metropolitan Cathedral
b		Palace of Fine Arts
c		Torre Latinoamericana

Q1. Look at the picture. If you are the person taking the picture, a/b/c is on your ______.
(Choose either one answer or two answers, one A/B and one C/D)

A. Left
B. Right
C. Front
D. Back

a	
b	
c	

Q2. Choose all correct answers.

A. b is ~200m north of c.
B. b is ~13min walk west of a.
C. To walk from c to a, one can walk ~1000m east and then ~400m north.

Virtual Scavenger Hunt (find this architecture):

C. Feedback Survey Questions in the Revised Experiment

Have you used each of the representation modes (the 3D mode and the Street View mode) before this experiment?

What are the key advantages of the 3D mode you found in helping you perform the wayfinding tasks?

What are the key disadvantages of the 3D mode you found in helping you perform the wayfinding tasks? Do you have any suggestions how it can be improved?

Do you want to use the 3D mode of Google Maps again? If so, for what purpose?

What are the key advantages of the Street View mode you found in helping you perform the wayfinding tasks?

What are the key disadvantages of the Street View mode you found in helping you perform the wayfinding tasks? Do you have any suggestions how it can be improved?

Do you want to use the Street View mode of Google Maps again? If so, for what purpose?

Imagine you are going to a place you have never been to before in 2 days and you will not be able to have access to any sort of maps or information about the place when you are there. Now you are given a chance to explore in either the 3D mode or the Street View mode of the place before you go. Which would you use? Why?