Trust in Adaptive Automation in a Tactical Search and Navigation Task

by

Anthony W. Broll

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

Master of Science in Aeronautics and Astronautics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2016

Signature redacted

Author ....
Department of Aeronautics and Astronautics
May 19, 2016

Certified by.. Signature redacted ....
Leia Stirling
Assistant Professor
Thesis Supervisor

Certified by.... Signature redacted .
Meredith Cunha
Draper Technical Staff
Thesis Supervisor

Accepted by Signature redacted ....
Paulo C. Lozano
Associate Professor of Aeronautics and Astronautics
Chair, Graduate Program Committee

Trust in Adaptive Automation in a Tactical Search and Navigation Task

by

Anthony W. Broll

Submitted to the Department of Aeronautics and Astronautics
on May 19, 2016, in partial fulfillment of the
requirements for the degree of
Master of Science in Aeronautics and Astronautics

Abstract

Handheld smart phone devices incorporate communication and mapping tools into small military squads to increase their effectiveness. These devices link a squad with headquarters, satellites and unmanned aerial vehicles to provide them with up to date intelligence. This information is filtered by adaptive technology which prioritizes the most vital pieces. An indoor search and navigation experiment was conducted to determine the appropriate type of automation (adaptive or adaptable) to prioritize this intelligence for decision makers in an uncertain, time-critical scenario.

An experiment was conducted with eight males in their early 20's actively serving the US military or part of a training program. Subjects utilized an app on an HTC Desire designed to navigate the user indoors from a start QR code to a goal QR code while collecting additional QR codes to maximize their point totals within three minutes. Subjects utilized the app in one of two modes: computer-select (adaptive automation) and user-choice (adaptable automation). In addition, updates in the form of floor closures would occur in half of the 24 trials. Results of the study showed a preference for computer-select with better performance on the primary task.

Users ended up using both systems as a type of user-choice by disregarding the app's path planning beyond the initial route. The user preference for computer-select was tied to the ability of the system to adjust to the human instead of the user having to tell the system what it was doing. Subjects wanted the flexibility of adaptable and user-choice before the trial to plan and define their own route, but once the trial began, the subject's temporal demand was too high to want to maintain that level of control beyond the subject's actions in the real world.

Thesis Supervisor: Leia Stirling
Title: Assistant Professor

Thesis Supervisor: Meredith Cunha
Title: Draper Technical Staff

Acknowledgments

This is the acknowledgments section. You should replace this with your own acknowledgments.

This thesis would have not been possible without the collective support and assistance of countless people. In a effort that cannot show my immense gratitude, I'd like to acknowledge a few here. First and foremost are my parents, Jane and Pete Broll, who have mentored, supported and loved me for the past 24+ years. From driving my siblings and I to school, coaching my soccer teams and putting a delicious homemade meal on the table (regardless of the hour), I would not be here without you. Next my siblings (Matt, Sam and Jess) for being the most competitive, sneaky and talented people a brother could want (sometimes). Thank you for keeping me grounded and being there for the countless moves and car rides. Next my grandparents, Grandma, Grandpa, Mo and Poppy for being such good examples for me as I grew up. The rest of my extended family for the chaotic family gatherings that remind me of the important things in life. My advisors, Leia and Meredith, for giving me the freedom to design my own project and mentoring me through the ups and downs over the past two years that have brought me to this point. My amazing UROPs, Matt, Solan and Nick. From the implementation of the app and WPS (Matt), to the running of the experiment (Solan and Nick), there is no way I could have done this without each of you. Dr. Alan Natapoff for the assistance in deciphering the arts of statistics and experimental design. My subjects, whom I cannot name, but I could not have graduated without. MIT, Draper and the Navy for giving me this opportunity. Finally to my MVL, Grim/sMITe and Boston/Cambridge friends, for making the past two years the best "vacation" ever.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter will present the motivation of this thesis with relevant literature review and the organization of the full thesis.

1.1 Motivation

The Department of Defense and DARPA established an initiative in 2004 to improve the combat effectiveness of each soldier by improving attention, working memory, workload, and executive function[39]. This initiative was designed to deal with the growing problems of friend, foe, and neutral identification, fratricide and collateral damage, while avoiding casualties and still accomplishing the mission objective[4, 30, 8]. The program's solution was to improve situation awareness (SA), as defined by Endsley (1993)[8], by monitoring vital signs, sifting through updated information, and better connecting units. The increase in SA would provide access to more information, which would in turn hopefully improve the likelihood of making the best decision in time sensitive environments[39, 41]. One of the ways this has been implemented is the integration of handheld smart phone devices[23, 41, 58]. These devices incorporate communication and mapping tools into small military squads to increase their effectiveness. These devices continually link a squad with headquarters, satellites and unmanned aerial vehicles (UAVs) to provide them with up to date intelligence[23, 41, 58]. This information is filtered by adaptive technology, which actively shifts the amount of system involvement of the user based on the workload and capabilities of the human operator and the specific situation[28]. For this type of scenario, adaptive technology is utilized to prioritize the most vital pieces of information and push them to the top. This intelligence is then available to decision makers in these time-sensitive, dynamically changing environments, such as the decision to advance forward in a firefight. Of course, these types of situations are not limited to the military world; all dynamically changing, time-critical environments can benefit from this type of technology, like emergency responders for example. The interactions between the human being and the adaptive device in these complex, tactical environment has not been well defined. The focus of this thesis is the understanding of the human operator and his or her use, interaction and needs with adaptive mobile devices in the context of trust, workload and SA. An indoor search and navigation experiment was conducted to determine the appropriate type of automation (adaptive or adaptable) in an uncertain, time-critical scenario.

1.2 Human/Automation Interaction

As computer-technology has developed, the role of computers in everyday life has expanded. An increasing number of tasks have transitioned from humans to computers. This transition has been applied through the implementation of automation, which is any "technology that actively selects data, transforms information, makes decisions, or controls processes"[28]. Sheridan proposed levels of automation (LOA) to describe the role of the human and computer in a system[45]. They range from LOA1 in which there is full manual control to LOA10 which is full autonomous control[45, 22, 34]. Every LOA in between has a varying balance of human and computer control. For an example of a system with multiple levels, in a car, there are parts run by the computer like automatic gear shifting or cruise control and those run by humans like turning and braking. If the system requires varying levels of automation due to changing workload states, then dynamic task allocation can be implemented, which assigns whether the human or the computer is in control of that particular task. Dynamic task allocation can be done by the human (adaptable) or the computer (adaptive)[43, 24]. Adaptable automation is the state where the human is the ultimate decision authority on the allocation of automation to specific functions in the system. A driver assigning cruise control highlights adaptable automation, while adaptive automation is when the computer is the decision authority. A car automatically engaging its brakes to prevent a collision is an example of adaptive automation. No matter which level is implemented, the automation's goal is to mitigate human error.

Automation is highly effective at time consuming repetitive tasks, but it also creates a new set of problems[28, 22]. These new issues include: changing the nature of the task, changing or adding cognitive demands, relegating operators to supervisory roles, causing confusion in the operator, and failing to account for the human's role[22]. With the introduction of automation, the system becomes the human and the automation together. This combination results in the sharing of the task. The changing of the task leads to different or increased cognitive demands and less involvement of the user which could cause confusion. In addition, the designer of the automation, may fail to account for the appropriate inclusion of the human into the new task. The integration of adaptive technology into military units is to provide the user with the maximum amount of information they can handle in a given scenario and reduce or increase that information as the tempo of the scenario adjusts. For example, Johnson et al. (2002)[21] conducted an experiment in which subjects were required to create flight plans within mission constraints of threat, arrival time and fuel while minimizing route cost. The fuel, arrival time and threat constraints changed dynamically once in the middle of the trial and automation was provided in three of the four conditions to replan the route within a set time-constraint[21]. The paper showed that in time-critical scenarios, full automation performed better than the two forms of partial automation and no automation, and recommended the implementation adaptive automation to balance time pressure and user involvement in the system[21].

"Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future"[8]. The key point here is that situation awareness applies to the environment or everything beyond a person's immediate task. Studies on pilots in air-to-air combat scenarios have shown that certain information is primary under specific circumstances but can be relegated to secondary tasks[8]. Secondary tasks that require parts of a subject's SA are used to determine their workload based on their spare attention from the primary task[20]. Dynamic task allocation provides this capability to manage workload and engagement. The goal is to maintain the workers involvement in an environment so that they have enough tasks to remain engaged without becoming overloaded, while also reducing boredom during lulls[43, 20].

Several studies have shown across diverse fields as pilots and factory operators that adaptable automation is preferred and is easier to implement[43, 20, 24]. For example, Sauer et al. (2012)[43] showed that there was no noticeable difference between the use of adaptive and adaptable automation in performance of managing a simulated life support system, but the advantages of reduced development costs (time and money) and increased active management supported the use of adaptable automation. Even in studies where the participant had comparable performance between adaptable and adaptive, adaptable was preferred due to higher confidence, trust and engagement in that mode[24, 43]. Adaptive automation has however been more supported by the literature to reduce cognitive overhead[22, 24, 48] even though it is more complex and therefore a more expensive system to develop and implement than adaptable automation[43]. Adaptive automation can be triggered by events, performance, workload and sensory information[43]. The sensitivity of these triggers may increase stress and workload of the operator as the system reacts to the participants' state and may not anticipate the current status in time thus placing additional stress on the participant or not provide a high enough workload[43].

1.3 Trust

Trust is the "attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability"[28]. As previously discussed, the implementation of automation to resolve one issue may create other problems such as operator confusion, increased cognitive demand, and changes in the user's job. These issues significantly impact a user's trust of the machine system. A strong example of machine-user distrust is exemplified by Microsoft's Office Assistant better known as Clippy. Clippy failed to follow human rules of etiquette, was inappropriate for most users' skill levels, was not able to develop the user's skills and then remove himself, and was not applied in an appropriate situation of need[50]. Clippy elicited strong amounts of anger and distrust from users[50].

Trust is not only vital in human-human relationships, but also in human-machine relationships. Trust in the system is dependent on the purpose of the machine, the process the machine uses to accomplish the task, and the actual performance of the machine[28]. Appropriate trust leads to effective use of the system, but overreliance (misuse), under reliance (disuse) and inappropriate reliance (abuse) lead to ineffective use of the automation[33]. A user's trust in automation guides reliance and utilization of the system when the complexity of the automation makes a complete understanding impractical, or when the situation demands fast-paced adaptive actions that procedures cannot guide[28].

The formulation of trust between automation and humans begins by the user relying on the assigned purpose of the technology until the user has a performance history with the project and develops an understanding of the process through operational and training use[28]. The feedback to the user from the system through use is processed through analytic, analogical and affective means. Analytic processes are the application of pure logic to assess the expected outcome of using the system as a guide for trust[28]. Analogical processes are the development of trust based on associations with characteristics of the system and its environment[28]. Affective processes dictate the trust between the user and system based on the emotions of both parties which in this case is just the user[28]. Affective has the largest impact upon user trust because it is used to filter the analytic and analogical inputs[28]. Affective also provides the signals for changes in trust in both directions[28]. Examples of each process to assign trust would be the use of an instruction manual for analytic, the assessment of an online review for analogic and the frustration of the user due to system failures for affective.

Studies have shown that reliable systems allow for users to maintain initially established strategies because the system has not deviated from their expectations[27]. The introduction of faults into the system causes the user's trust to change and thus they change their strategy to find a new system reliability[27]. For example, in Lee& Moray (1992)[27] subjects used a mixture of automated and manual control to operate a orange juice pasteurization plant. In this study, most participants determined that the task was simple enough to use manual control and it produced better results than the automation initially, but when failures began occurring, the automation was used to reduce workload and attempt to find consistency among the other controls[27]. Based on Lee& See's model of trust, Miller (2005)[31] proposed that the systems' etiquette appeals to analogic and affective processes within trust and that this etiquette may have a great effect on tuning trust within the adaptive-automated relationships with humans. As shown by Clippy, systems should obey human rules of etiquette, should be applied appropriately to a task that fits with the user's experience, should develop user's skills to become self-sufficient, should be branded appropriately and should only be applied to a situation with a real need[50].

Within any experiment, multiple factors can have a positive or negative effect on the user's trust, to include complexity of task, training, stressors (time constraints, noise, sensory overload, fatigue), reliability, communication, workload, attitude, self-confidence, automation states, risk and predictability. The multitude of events and sources that effect trust make measurement very difficult. Attempts have been made to quantify trust in experimental setting[52] by relating system uncertainty to subjective trust of the system along the four metrics, of competence, predictability, reliability, and persistence, with only competence showing the only relationship. Unless uncertainty is clearly defined within the experiment and controlled, evaluations of trust must rely upon Likert items[52]. Jackson et al. (2015)[16] defined trust of autonomy based on seven heuristics which included: visibility of current and probable system behavior, accessibility of system rationale, visibility of system capabilities and limitations, awareness of latency and delays, transparency of failure, and fit with users and operations. Overall though, experiments utilize Likert items to define a Likert scale based on similar heuristics to measure trust[16, 55, 52, 19, 9, 51]. Generally, humans invariably trust themselves more than the system and attempt to maintain as much control as they believe they can handle regardless of whether it leads to better performance[43].

There have been numerous studies that have looked at trust and automation[27, 28, 31], and the differences between adaptive and adaptable automation[43, 24]. These studies have principally looked at LOA9, supervisory control[45]. An example of supervisory control is a factory plant operator. In LOA9, the operator is asked to put in initial values and monitor the system, while the system is asked to take in new information from the sensors, to perceive the importance of the information, to decide what action to take, and then to act. The operator is only asked to step in if an emergency occurs. This can lead to out-of-the-loop syndrome, where the user lacks the awareness of the situation and the system to be able to step in appropriately[20]. This level of automation is not suitable for all scenarios and adaptation in the form of adaptive or adaptable automation may benefit users more effectively. In a tactical military scenario for example, the operator is using an automated hand held smart phone that is providing communication, SA and recommended actions. Unlike the factory worker, the smart phone can only take in the information, consolidate the information and make a decision of what action the operator should take. The smart phone cannot perform the action. The operator must refer and interact with the device to fill in the information gaps vital to their decision while maintaining their SA. If the operator does not like the recommendation, the operator can choose to ignore the system. This type of interaction increases the importance that trust plays in the human-machine dynamic because a lack of trust could result in disuse and too much trust could result in over-reliance. With active-moving soldiers and constantly changing environments, a tactical military scenario adds more dynamics, complexity and factors than a supervisory control setting. In addition few studies have directly compared the effects of adaptive and adaptable automation on human and system performance[24]. This thesis seeks to extend the literature by comparing automation types in a mobile environment.

1.4 Selection of Experiment Environment

This thesis expands understanding of the balance between adaptable and adaptive automation in a tactical environment. Here potential environments include: outdoor, indoor, video game and virtual reality with a treadmill. A review of literature was conducted to look at the use of virtual reality (VR) and video games as possible environments. Huberman & Glance (1993)[14] showed that the randomness and uncertainty of the real world can be lost in computer simulations when synchronous updates and precise feedback are included. This demonstrated the impact of program designers on the computer environment utilized by a participant. Witmer et al. (1996)[56] showed that the use of VR in a training environment for the transfer of route knowledge produced better results than the use of only manuals and lectures, but it is also limited by programming realism. Table 1.1 summarizes the tradeoffs for the environments in consideration.

The main requirements for the decision were the desire for a realistic environment that would require physical and mental demands, the ability to conduct the study in the weather of the Boston/Cambridge area and the variety necessary in course setups to keep the subject engaged for up to 30 trials. The motivation for realism ruled out video games and VR, while the outdoor environment was removed due to weather and variety concerns. Thus the selection was narrowed down to a tactical indoor environment. Buildings 35 and 37 at MIT were chosen for the proximity to the Man Vehicle Lab and AeroAstro.

Table 1.1: Possible Test Areas

	Tactical Indoor	Tactical Outdoor	Video Game	VR + Treadmill
Pros	•Control for weather •Dynamic Environment •Physically and cognitively demanding •Real World Environment •Variety through Floors •Small Space •Course Setup	•Dynamic Environment •Physically and cognitively demanding •Real World Environment •GPS measurements	•Environment Variety •Cognitive Demands •Control for weather •Positioning of objects and subject preprogrammed •Low logistic demands	•Environment Variety •Immersed in the system •Control for weather •Dynamic Environment •Physically and Cognitively Demanding •Programmed positions and objects •Easier to do physiological measurements •Low logistic demands
Cons	•Variety of environment •Interference of other humans •Need more assistance in running experiment •New programming side of building •Sensor Concerns for subjects	•Variety of environment •Interference of other humans •Need more assistance in running experiment •Weather Concerns •Large Scale	•Programming the game •Not real world •No physical demands •No immersion •Not dynamic environment	•Not real world •Always running in the same direction •Always have same speed •New programming for combining Oculus Rift with navigation system •Weight on head

1.5 Algorithm

To enable a real time navigation system, a path planning algorithm was required. The algorithms in Table 1.2 were explored as possible solutions to the problem. The pros, cons and relationships are shown in the table. The path planning algorithm was intended to be applied to a graph system of nodes connected by edges. The nodes represent specific locations and the edges represent the connections between them. For example a node on a floor could have connections with other nodes in a hallway or up a staircase. The requirements for the path planning algorithm were that it be simple to implement, able to find the shortest path between two locations, produce paths able to prioritize multiple locations, allow varying edge costs to account for floor changes, be computationally time efficient, create multiple paths and allow doubling back. Based on these criteria, a perfect algorithm was not found. A* is simple to implement, can produce the shortest path between two nodes, allows for varying edge costs and is computationally time efficient. As the closest option, A* was selected as the basis of the path planning algorithm with modifications to allow for multiple priorities, multiple routes and doubling back. The solution actually implemented will

Table 1.2: Possible Path Planning Algorithms[6, 7, 12, 59, 35]

Algorithm	Feature Pros	Feature Cons	Builds Off
Breadth 1st Search (Cormen et al. 2009; Correll 2014)	• Simple to implement	• Checks every node • Uniform edge costs • Does not allow doubling back • Only produces one path	
Dijkstra (Cormen et al. 2009; Correll 2014)	• Simple to implement • Allows edge costs • Shortest path	• Checks every node • Time intensive • Does not allow doubling back • Only produces one path	• Breadth 1st Search • Greedy
Greedy (Cormen et al. 2009)	• Simple to implement • Does not check every node • Allows edge costs • Computationally time efficient	• Not optimal • Does not allow doubling back • Only produces one path	
A* (Hart et al. 1968; Correll 2014)	• Simple to implement • Does not check every node • Shortest path • Allows edge costs • Uses heuristic • Computationally time efficient	• Does not allow doubling back • Only produces one path	• Dijkstra
Yen or K* (Yen 1971)	• Produces several paths • Based on shortest path • Allows edge costs	• Computationally time intensive • Does not allow doubling back • Not every path is useful	• Any shortest path algorithm (Dijkstra and A* for example)
Travelling Salesman Problem (Correll 2014)	• Visits every node • Allows doubling back • Allows edge costs • Allows multiple priorities	• Visits every node • Computationally time intensive • Only produces one path • Inefficient route	

be detailed in Chapter 2.

1.6 Indoor Positioning System

Designing a mobile navigation and search task for use in an indoors test environment required an indoor positioning system (IPS). IPS has been a very popular area of business research. Retail companies are looking at ways to track user interest and offer them other items related at their current location[60]. Using Bluetooth and WIFI, Google has been able to implement an IPS in campuses, stadiums and malls[60, 11]. Other technologies that have been used in IPS include echolocation[18], radio-frequency (RF) identifications (ID), ultrasound beacons[36], inertial systems, LEDs, and magnetic fields[44]. The use of multiple technologies alongside sensors such as gyroscopes, accelerometers, magnetometers and altimeters results in increased accuracy[1, 32].

Due to the frequency of WIFI access points and quality of WIFI strength within MIT's buildings and the lack of additional sensors being required, WIFI was selected as the primary source of positioning in this experiment. There are four methods for a WIFI IPS: cell ID, triangulation, trilateration, and fingerprinting[5]. Cell ID utilizes WIFI access points (AP) to determine position based on the nearest WIFI AP[5]. Cell ID's accuracy is heavily dependent upon the proximity of APs. Triangulation uses the given coordinates of APs and the angles between them and the user to triangulate a position[5]. Trilateration uses the distance from multiple APs to find the intersection point between those distances of at least three APs to determine the position[5]. Triangulation and trilateration rely on the full complement of APs with the system being operational at all times and precise definitions of all APs in the area of interest. Fingerprinting surveys the WIFI in a location to create a database of the signal strengths at a specific location within the known map and stores the information in a database[5]. At a later time, measured WIFI strengths are compared to the database to determine the user's current location[5]. Fingerprinting involves surveying, while the rest require input of WIFI APs[5]. Fingerprinting has been demonstrated to have within one room accuracy, but increased human traffic can reduce this accuracy to a little over 60 percent[5]. Fingerprinting relies on relative WIFI strengths and is therefore less dependent on the utilization of the full system architecture and therefore more flexible in the larger system[5].

While there are apps in the Google Play store for IPS (Indoor GPS[26], WIFI Compass[57], Build NGO - Indoor Navigation[42], Crux Indoor Location[47], and Infsoft Maps[15]), the capabilities seemed limited for modification. Thus an IPS based on WIFI fingerprinting was developed for the Android development environment, which had been shown to be successful[46, 29, 38]. Challenges were still expected with the implementation of a WIFI fingerprint IPS. With WIFI strengths not confined to specifics floors, floor determination has been shown to be challenging[38]. In addition extensive upkeep of the server would be required to ensure the accuracy of the data base through out the length of the experiment.

1.7 Sound Considerations

A tactical, mobile experiment that is changing in real time requires a way to direct user-attention. There are numerous ways to direct attention using auditory signals; this section will examine the use of sound earcon, spatial soundscape and spatial 3D soundscape. First a brief description is provided, followed by more detailed explanation. Table 1.3 provides an overview and summary of pros and cons.

Sound earcon is the utilization of a specific sound to signal an event or other piece of information. A spatial soundscape provides information about the user's proximity to an event or object to which the user is being directed[53]. Finally, a cue in a spatial 3D soundscape provides the user with the proximity and directionality in 3D space of the object of interest. It accomplishes this by changes in volume as it attempts to replicate a human's auditory system[53].

Sound earcon is the simplest of the options. Earcons are "audio messages used in the user-computer interface to provide information and feedback to the user about computer entities"[3]. Specialized types of sound earcons include: auditory icons, spearcons, lyricons and compound earcons. An auditory icons is the "utili[zation of] metaphors to relate the[user] to their virtual referents" for example "a tyre-skidding is used in vehicle collision warnings"[10]. Spearcons are the sped up version of text to the point where the word is no longer comprehensible[54]. A lyricon "combin[es] the two layers of musical speech sounds (lyrics) and non-speech sounds (earcons) concurrently"[17]. Finally, compound earcons are the connection of multiple earcons in a string[3]. In general sounds should not contain more than a four note phrase that varies rhythm, pitch, timbre and register dynamics. This briefness is designed to prevent melodic creations which could cause users annoyance[3].

A spatial soundscape provides proximity but not directionality to the triggered event. There are two major ways that this can occur: changes in frequency or changes in volume[53]. For changes in frequency, the earcon is repeated at smaller intervals as the user approaches the object of interest and longer intervals as the user moves farther away[53]. For changes in volume, the earcon would crescendo as the user

Table 1.3: Possible Sound Implementations

Sound Family	Sound Information	Pros	Cons
Earcon (Blattner et al. 1989; Grazonis et al. 2009; Jeon & Sun 2014; Walker et al. 2006)	Basic Earcon (Blattner et al. 1989)	• Simple	• No intuitive mapping
	Auditory Icon (Grazonis et al. 2009)	• Simple • Could provide intuitive mapping depending on situation	• Hard to find appropriate mapping metaphor
	Spearicon (Walker et al. 2006)	• Simple • Could provide intuitive mapping depending on situation	• Must learn mapping
	Lyricon (Jeon & Sun 2014)	• Simple	• Must learn mapping • Melodic could incite irritation by user
	Compound Earcon (Blattner et al. 1989)	• Simple	• No intuitive mapping • Unnecessary unless several sounds are required
Spatial Soundscape	Spatial Soundscape (Vasquez et al. 2012)	• Provides proximity information	• Complex • Requires accurate positioning
Spatial 3D Soundscape	Spatial 3D Soundscape (Vasquez et al. 2012)	• Provides proximity information • Provides directionality	• Complex • Requires knowledge of head orientation • Requires accurate positioning • Confusion between backward and forward directions • Requires headphones

approached the object and decrescendo as the user left the area[53]. Spatial 3D soundscape provides the same information as a spatial soundscape with the addition of directionality[53]. Table 1.3 was formed based on the literature and summarizes the pros and cons of each possible sound implementation.

Based on studies, earcons generally performed worse in studies to order objects, locate menu items and determine the correct relationship between a sound and a service than spatial soundscape and spatial 3D soundscape, and was simplest to implement[10, 17, 54]. Auditory Icon "performed significantly better in terms of intuitiveness, learnability, memorability and user preference" than basic earcons in a study conducted by Grazonis et al. (2009)[10] to determine the correct relationship between a sound and a service. Spearcon also performed significantly better than basic earcons and as well as if not better than auditory icons according to Walker et al. (2006)[54] in navigating between menu items. Lyricon is only in its initial study phases, but was shown to have higher accuracy rate in ordering subjects in a sorting task and reduced mapping time in subjects as compared to basic earcons[17].

A spatial soundscape was reported to be "jumpy" and "not proportional" with volume changes during one study where the soundscape relied on the use of GPS to initiate the changes[53]. This was most likely due to the systematic error associated with GPS[53]. Spatial 3D was chaotic if more than two sounds were overlapped, but overall performed better at immersing subjects into a garden exploration task than spatial and earcon by better directing and holding user's attention to landmarks within the garden[53].

For the scenario being considered, the only requirement was to be able to draw the user's attention to changes in the mobile device. With this requirement, the complexity of a spatial soundscape or a spatial 3D soundscape were unnecessary because proximity and directionality were not needed. In addition, there were concerns about having knowledge of the head orientation and accurate positioning of the user within the global map. It was determined that three sounds would be required, so a earcon would be appropriate. The goal was to find sounds with as useful mappings as possible. For this reason, auditory icons were selected for the scenario being considered. The selection will be discussed in Chapter 2.

1.8 Research Goals and Thesis Organization

The purpose of this thesis is the exploration of the effect of adaptable and adaptive automation on performance of the user in an active navigation and search task. This will provide insight into the effectiveness of automation in a tactical scenario, into the effect on the user's trust of the system and into the interaction between a mobile interface and a user in a scenario with time critical decision making. As previously discussed, military ground forces currently utilize smart phones to connect squad units with commands from headquarters, intelligence from UAVs, satellites and other units, and with each other. The challenge is to provide the user with the necessary information to make the best decision while not overloading the user.

A simple navigation and search task will be tested to learn about the extent that the user wants to be involved in the decision process in a time sensitive scenario, which may or may not change during the task. This research will help understand the interactions between operators, such as first responders and military personnel in the field, with handheld smart phones providing assistance. Specifically it will look at three questions:

1. Is there a difference in performance between adaptive and adaptable automation?
2. Is there a difference in performance with and without an update to the scenario?
3. Is there an interaction between automation type and update case for a performance measure?

Chapter 1 focused on the motivation of the thesis. Chapter 2 will cover the design of the app used for the experimental task. Chapter 3 will discuss the experiment completed by the subjects. Chapter 4 will present the results and discussion of the data collected. Finally, Chapter 5 will conclude on the contributions and limitations of this thesis and future work to explore.

Chapter 2

Application Design

As discussed in Chapter 1, few experiments have been designed to look at adaptive automation in mobile, tactical scenarios where the user is the primary actor in the computer-user system. Thus an experiment has been designed which will be discussed in Chapter 3. This chapter will explain the design of the app utilized on an HTC Desire during the mobile search and navigation task, where the user navigated an indoor environment to collect points. The three main functions of the app were to navigate the user in the hallways, assist the user in the collection of points and provide time keeping functions. In addition, the design of the seven unique scenarios will be discussed.

2.1 Path Planning Algorithm

To set the stage for the path planning algorithm, we will first define the environment in which it operated and then discuss the algorithms priorities and goals. As described in Chapter 1, each path planning algorithm was dependent upon a clearly defined connectivity within the map or graph. Figure 2-1 shows the hallway layout of buildings 35 and 37. The dark blue circles signify the nodes available on each floor. Each node's connections are shown by the dark blue line connections on each floor. Green dashed lines signify staircase connections and orange dashed lines connect the buildings to each other. Each node is assigned as either a start location, goal loca-

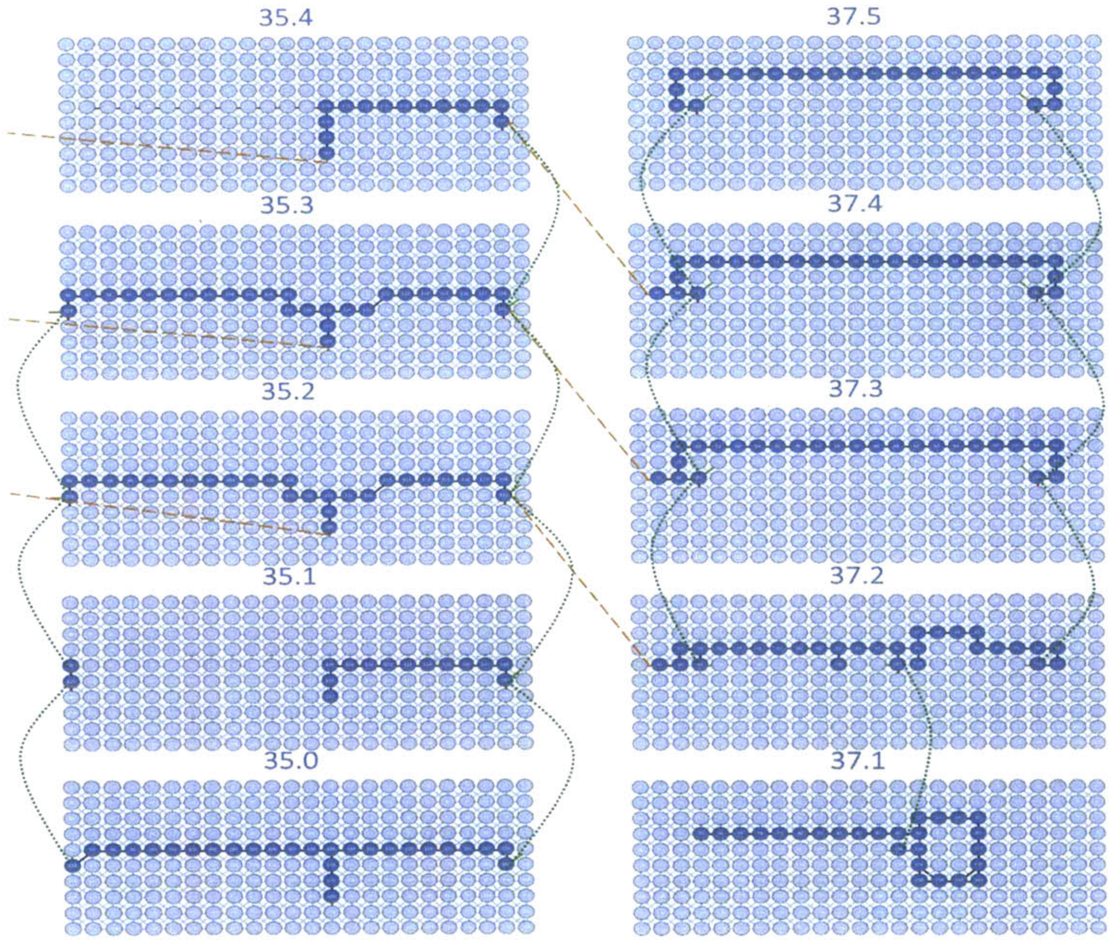


Figure 2-1: Grid overlay for MIT's buildings 35 and 37.

tion, QR code location, stair node location or basic node location. Every node was assigned a base weighting referred to as a floor weighting in the algorithm. In addition, each stair node was separately assigned with a higher weighting to represent the additional expense of transversing a full staircase. There was one case where the floor weighting was assigned a value larger than the stair weight. This was an alternate path provided in the user-choice mode, which will be described later in the chapter. Other than floor and stair weightings, time estimates for each node were assigned with differences provided between floor, stair and QR nodes. More detail will be provided on the assigning process later in the chapter.

Fundamentally the path planning algorithm originated as A* from source code on Google Code[49]. However it was quickly determined that a pure A* would not provide us with the precise solution we desired. A* is excellent at finding the shortest path from point A to point B, but this experiment did not necessarily desire that. The goals for the algorithm desired were to arrive at the goal node by the 180 second time limit and to maximize the points gathered by the user in that time. To meet these goals, the initial A* algorithm was modified in four key ways: gradient weightings for QR codes, removal of the list of previously visited nodes, a heuristic constrained by time estimates and node weightings, and creation of intermediary goals.

In order to encourage the collection of points, QR codes were regarded by the algorithm as having a weighting of the base floor minus the value of the QR code. The nodes in the vicinity were also reduced in a gradient effect with subsequent neighbors being worth one more cost value. This gradient was experienced by the four subsequent neighbor sets on either side of the node to hopefully entice the algorithm to head towards these nodes.

Second, the list of previously visited nodes was removed from the algorithm. The purpose of this list is to limit the subsequent search of future nodes and continue pushing the algorithm to the goal. In basic A*, the algorithm automatically discounts nodes that have already been selected and visited. The removal of the list of previously visited nodes allowed for the algorithm to double back on itself.

Third, the algorithm was constrained by a heuristic based on time estimates and node weightings associated with each node. The floor weightings were used to direct the algorithm to its current goal, while the time heuristic was used to constrain the algorithm from visiting locations that were too far from the final goal as the time expired.

Finally, the algorithm was modified to become more like a Greedy algorithm by having an intermediate goal list. This goal list was predefined during each scenario and was created by looking for the local QR code worth the most points. The local region was defined as all nodes within a cost of 120 with the general base floor cost being ten per node. This was referred to as the goal list window. Each goal list consisted of at a minimum four initial intermediate goals in addition to the ultimate goal of the end node. If the user deviated from the given path and goal list, a new goal list was created in real time from the current location. If there were no QR codes within the region, the search would expand by 20 until it had a goal to add to the list. This goal list was time sensitive during the experiment and, depending upon its preset time values, would break the goal list loop to reach the end by the three minute time limit. It accomplished this by assigning the only goal of the algorithm to be reaching the goal node.

In essence, the algorithm directed users on the shortest path to the next intermediary goal with possibly slight deviations to reach additional nearby points. The algorithm was not constrained by its previous visits and thus capable of doubling back. Overall, it was time constrained by reaching the final goal node. K* was used to validate the algorithm to make sure useful routes were created for the three routes provided to users by comparing them to k additional routes based on the A* shortest path[37]. TSP was used to make sure the goal lists were beneficial by comparing the local goal lists to the optimal path to collect every QR code on the map[2].

In order to select the inputs for computer-select and the additional routes for user-choice, map variations were examined with the following changes: base weight, stair weight, time for floor, stair, and QR nodes, goal list window, allowed time and break time. Allowed time defined the time for a route that a user would be willing to attempt, for example the subject could know that they had 180 seconds, but thought they would be able to shave 30 seconds off the time and attempt a 210 second route. Based on analysis of the proposed routes, allowed time was eventually held constant at 180 seconds. Break time was the time from the end when the algorithm would leave the goal list and direct its full attention to final goal, to ensure the user would arrive on time. Node time estimates were determined by timed walk-throughs of the building in segments. These can be seen in Figure 2-2. The full scenario assignments can be seen in Appendix A.

These estimates led to expected values of approximately two to three seconds per floor node, three to four seconds per QR code node and 10 to 15 seconds per stairway. These values were used in the route creation loop in addition to the rest of the values in Table 2.1. The numbers in bold were the base values used for the computer-select

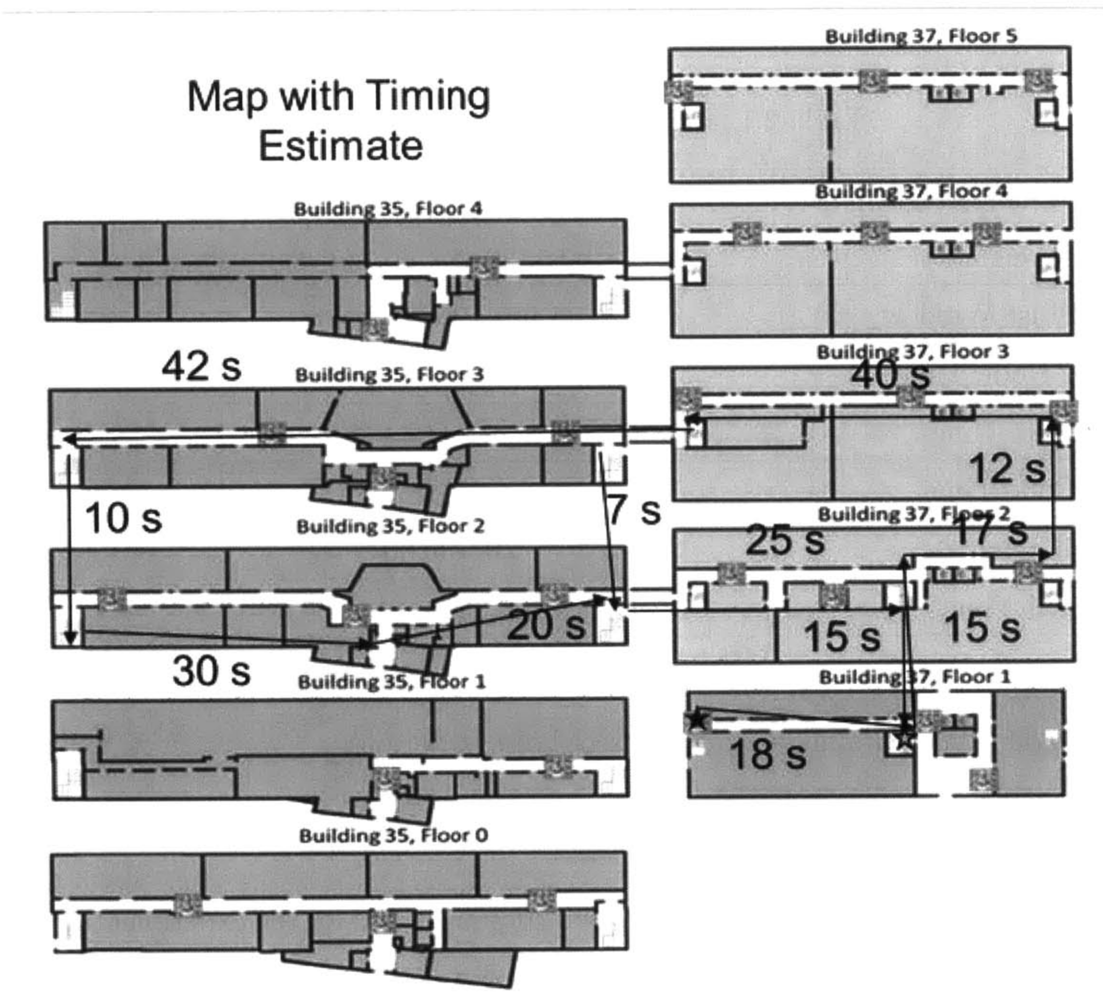


Figure 2-2: Map Showing Time Estimations

mode. The values italicized were selected in various user-choice routes. The rest of the values were eventually not selected. The selected values for the computer-select routes were chosen to allow for plenty of time for the subjects to orient themselves during the experiment and collect exactly three QR codes. The two user-choice variations added were chosen by comparing points collected, uniqueness and arrival time estimate. The route parameters were not consistent between options and trial scenarios. Scenarios will be explained in more detail in section 2.2.

Table 2.1: Route Selection

Base Floor Weight	**10**, *20*
Stair Weight	**15**, 20, *30*, 40, *50*
Goal List Window	90, **120**, *150*
Floor Node Time	*2*, **3**
Stairway Time	**10**, *15*
QR Code Time	*3*, **4**
Break Time	**20**, *50*, 80
Allowed Time	150, **180**, 210

Bold base values for computer-select

Italics additional parameters for the user-choice variations

2.2 Trials Selected

For this experiment, seven unique trials were designed. Six for the experimental trials and one for training. Those trials were designed with three thoughts in mind. First there was to be no repetition of start locations or goal locations, since the user would repeat trials. Second, the start and goal locations had to be no farther than 45 nodes apart which was approximately 90 seconds by the time estimate. Finally, each trial would have a point spread of one QR code for point values 4-9 and two QR codes for point values 1-3. This meant that in a given trial, 14 of the 28 nodes were active as either a start, goal, or point location. Figure 2-3 shows the generic map with all of the QR code locations and floors utilized.

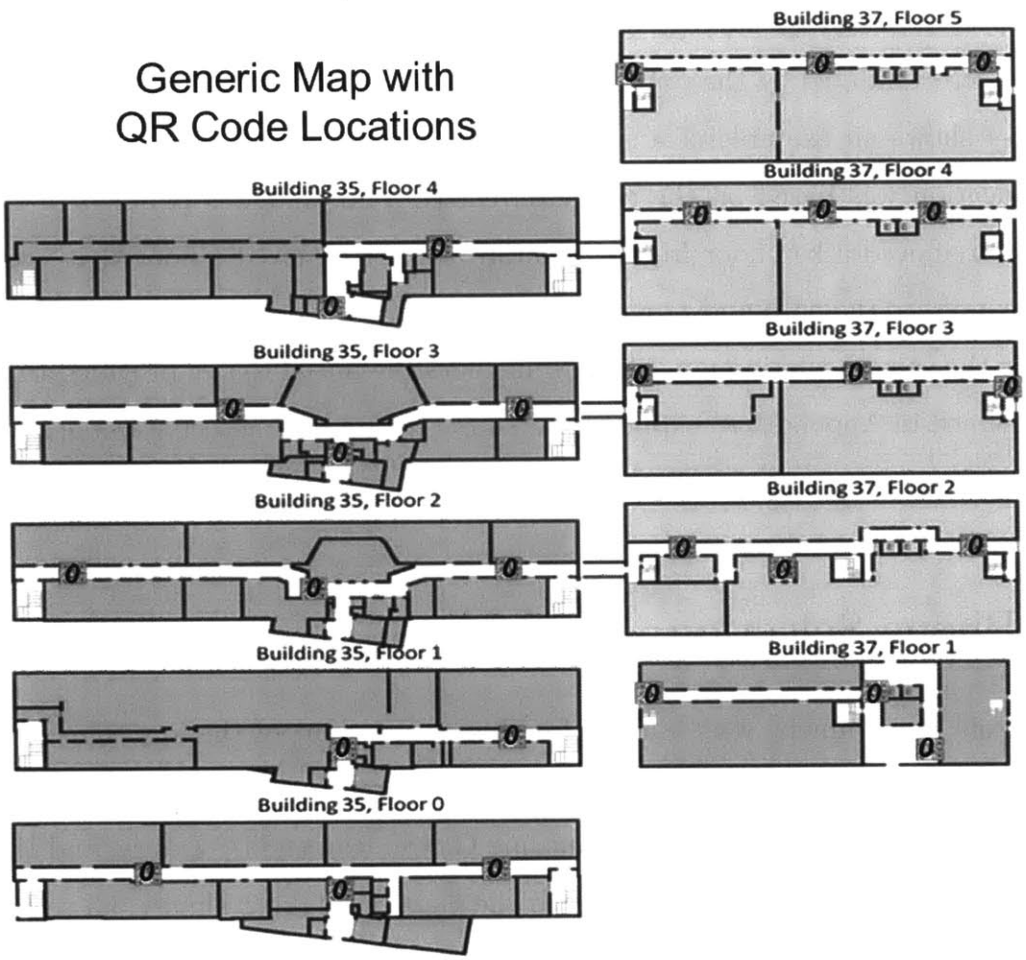


Figure 2-3: QR Code Locations

Appendix B, shows the six trials utilized throughout the experiment and the training trial. They are labeled with the NATO phonetic alphabet, with the first letter being the trial without an update and the second trial having the update listed in the bottom right-hand corner of the figure. Each trial was designed to create interesting routes for the user and provided three unique routes based on different inputs to the algorithm as mentioned in the previous section.

Points were placed by hand to create multiple path options for the user. These locations were validated by the route selection process and experimenter verification. Figure 2-4 shows an example of a scenario used during the experiment. The floor closure location was based on the computer-choice route and was planned to be a route that removed a key floor from the initial plan, or removed a floor that was in close proximity to the path and provided possible points to add to their route. Figure 2-5 shows the same scenario map, but now includes the initial routes provided to the user as defined in Appendix A. Appendix C shows the remaining scenario maps with initial routes.

2.3 Phone Selection

The Android environment was selected for the development of the app, due to accessibility and ease of development. From this requirement a phone was selected. Evaluation of the Samsung Galaxy 4, Samsung Galaxy TabS, HTC Desire, and HTC Eye revealed varying WiFi scan rates, with the Galaxy 4 taking almost 3.5 seconds and the HTC Desire taking 500 ms. Due to these variations, the HTC Desire was selected to allow for more reliable and quicker map updates. This phone was equipped with a magnetometer and three accelerometers, but no gyroscope. Therefore there was no knowledge being provided of the phone's orientation in the physical world. This prevented the collection of data to determine actual velocity and acceleration of participants during the study. While speed penalties were set in place to protect students, faculty and staff, there was no way to specifically determine the actual speed. Thus speed penalties were to be enforced by the experimenter, but none were given in

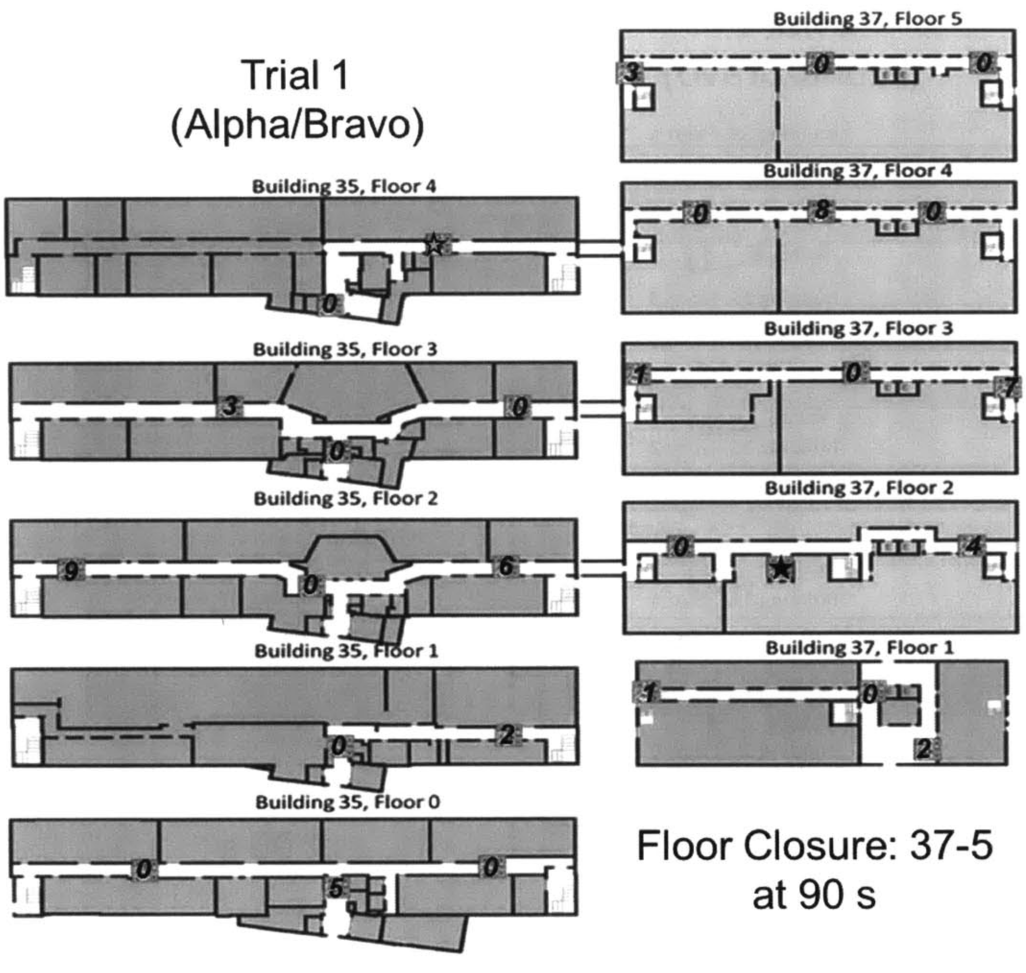


Figure 2-4: Trial 1 Scenario Map

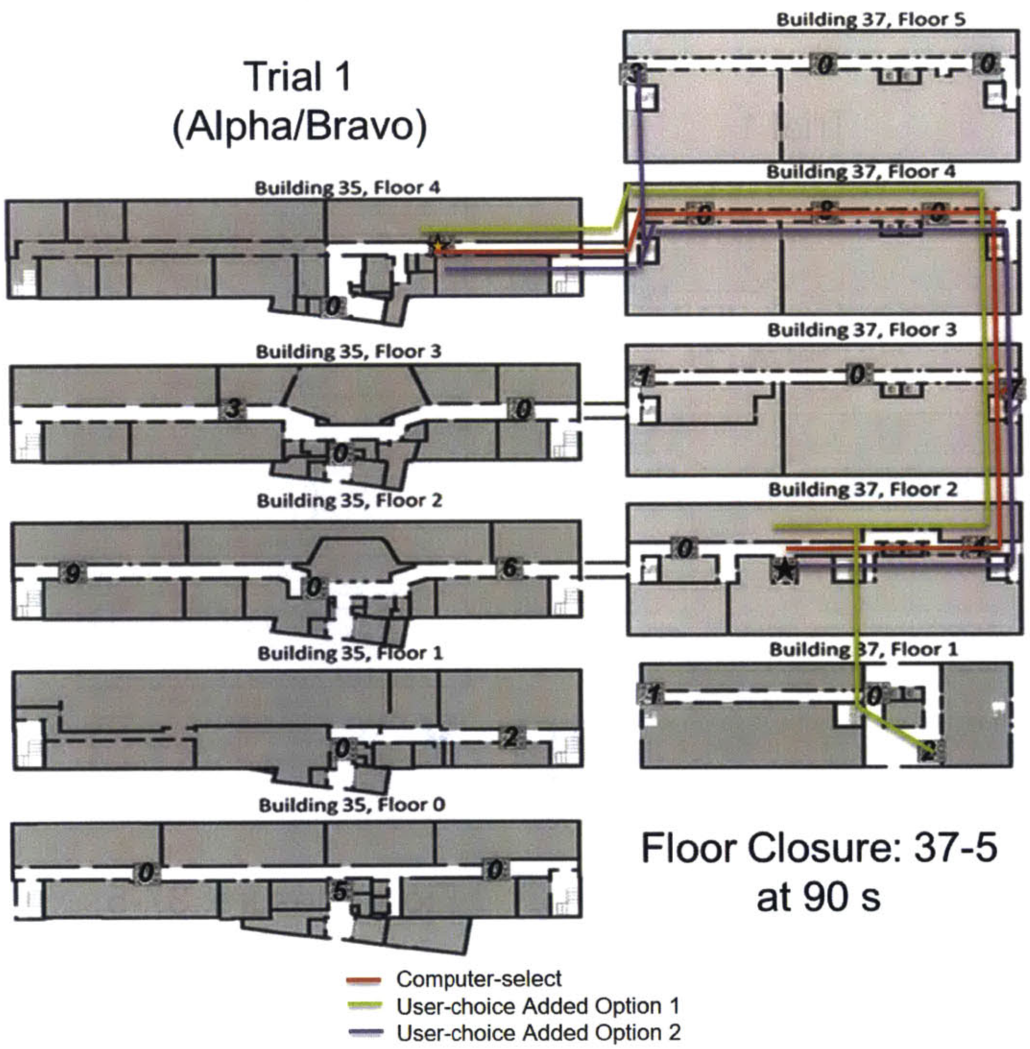


Figure 2-5: Trial 1 Scenario Map with Initial Route Options

this experiment. The power button was blocked off to prevent the app from resetting in the middle of a trial.

2.4 Overview of App

The study, which will be described in more detail in Chapter 3, took place over three separate days, with one session each day. The first session was for training while the remaining two were for testing. Over the two testing days, there were 24 trials. During each trial, the subject utilized a phone-based navigation aid. The aid provided them with maps (see 2.6), tracked their location using a WiFi-positioning algorithm (see Chapter 1), provided sound alerts (see Chapter 1) and provided them with a recommended route for each trial (see Chapter 2).

At the beginning of each trial, the subject was assigned one of the two automation modes (computer-selected or user-choice: see below), a start and end location, and a time-limit of three minutes to reach the desired end destination. The scenario space covered two buildings with 10 different floors and five different stairwells. While navigating throughout the hallways and stairwells the subject collected points by scanning QR codes in the hallways. The primary goal was to collect the maximum number of points while reaching the end location in the time allowed. The secondary goal was to fulfill a secondary situation awareness (see Chapter 3). In addition to the two modes, during half of the trials the subject received updates that closed a specific floor.

Throughout the rest of the chapter, the design of the app will be discussed. Specifically looking at the navigation modes (2.5), views (2.6), and features (2.7) of the app.

2.5 Navigation Modes

The two different navigation modes were computer-selected and user-choice. The subject used one or the other for the totality of a trial as defined by their treatment (see Chapter 3). In computer-select mode, the app provided all route planning functions without input from the user beyond their physical movements. While in user-choice mode, the app required the user to select one of three planned routes to traverse based on points available, number of staircases and estimated time. These will be discussed below. In addition there was a walk-through mode utilized during training.

2.5.1 Computer-Select

In computer-select mode, the user was given a navigation route to maximize points and reach the goal QR code in time. The computer-selected route was dynamic and incorporated any user deviations within approximately 10 meters. For example if the user thought they could gather more points by deviating off the given route, the app would provide a new route to account for the user's deviation from the path in real-time. Depending on the deviation, the new route may have the user return to their old route, or recommend a completely new path. Rerouting could occur for reasons beyond deviation to include: slight end of route modifications based on the node that the break time initiated, the user moving too slow, the inclusion of an added QR code because the user is ahead of schedule and a floor closure.

However if the subject did wander from the route a significant distance, for example to a different floor than was included on the route, then the navigation system would calculate a new route for them. If a new navigation route was calculated, a sound notification occurred and an alert window popped up.

2.5.2 User-Choice

In user-choice mode, the subject was given three navigation routes to pick from to maximize points and reach the goal QR code in time. The subject selected a route from three options at the start of the trial and then could change the route to follow at any time during the trial through the Alert Bar (2.4.1). These options can be seen in Appendix C for the initial routes.

Similar to computer-selected, the user-choice navigation route would recommend a modified route for any of the possible rerouting reasons listed in the previous section.

However, the navigation route was not dynamic as in the computer-selected mode. If the subject wandered from the route a significant distance, then they were prompted to decide whether to choose a new route or not. If they wandered too far off path, then the current route could be too far to reach the goal in time. Refusing a new shorter route recalculation could result in failure to successfully complete the trial within 180 seconds. If the subject decided to not choose a new route, they were not prompted again for the next 10 seconds regardless of their deviation.

2.5.3 Walk-through Mode

In addition to user-choice and computer-select modes, there was a walk-through mode developed for training. This mode provided no route planning. The purpose was to provide an environment for the subject to explore the two buildings being used during the experiment. The goals of this mode were to find the 28 QR code locations and five stairwells, while getting a feel for using the QR code scan feature (Section 2.7.2).

2.6 Map Interface

Regardless of mode, the user had two primary views that they could utilize of buildings 35 and 37. The two views were displayed in the app with two different perspectives, one from above called the Top View and the other from the side called the Side View. The Top View (Figure 2-6) shows the floor plan of the current floor from above, while the Side View (Figure 2-7) shows the global perspective of the two buildings.

2.6.1 Top View

The Top View shows the floor plan of the current floor from above. The current position is shown with a green circle on the map. Figure 2-8 shows the Top View with callouts.

The floor plan and map displays did rotate to align North nor to align with the phone orientation, instead they were static like using a paper map. A solid blue path

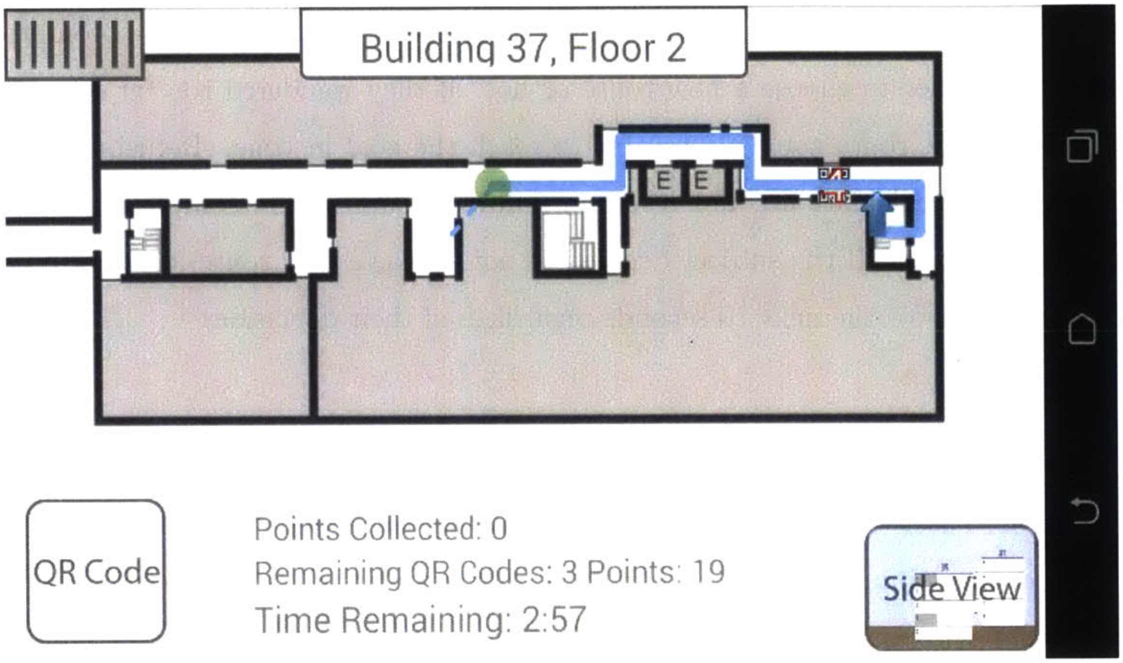


Figure 2-6: Top View

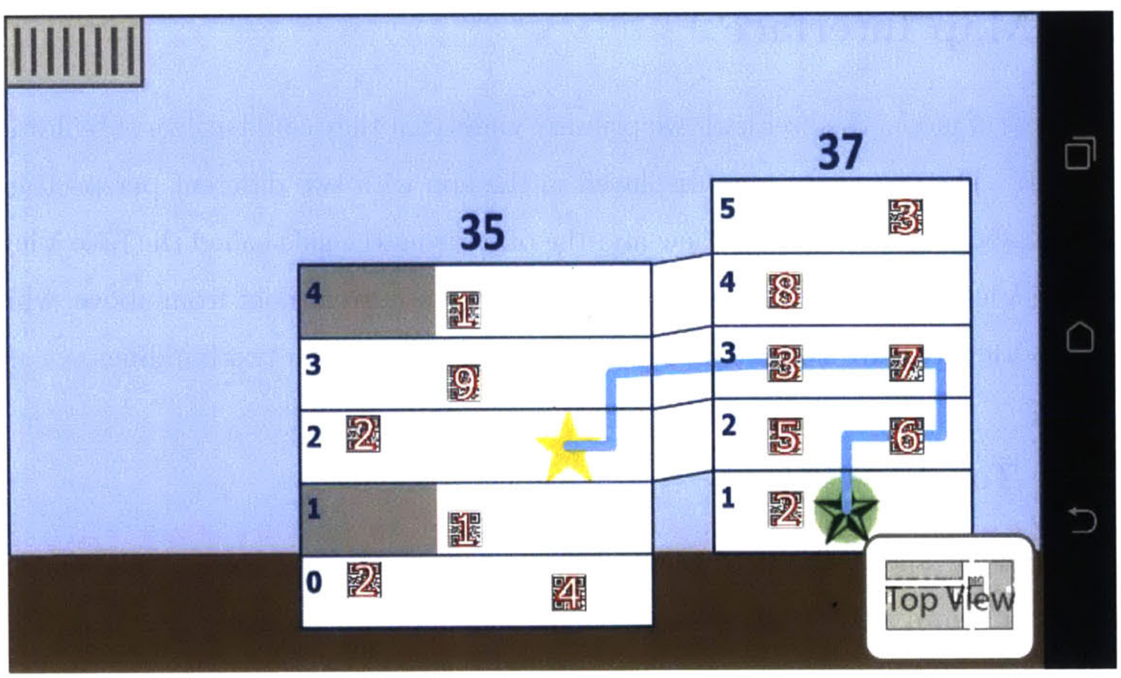


Figure 2-7: Side View

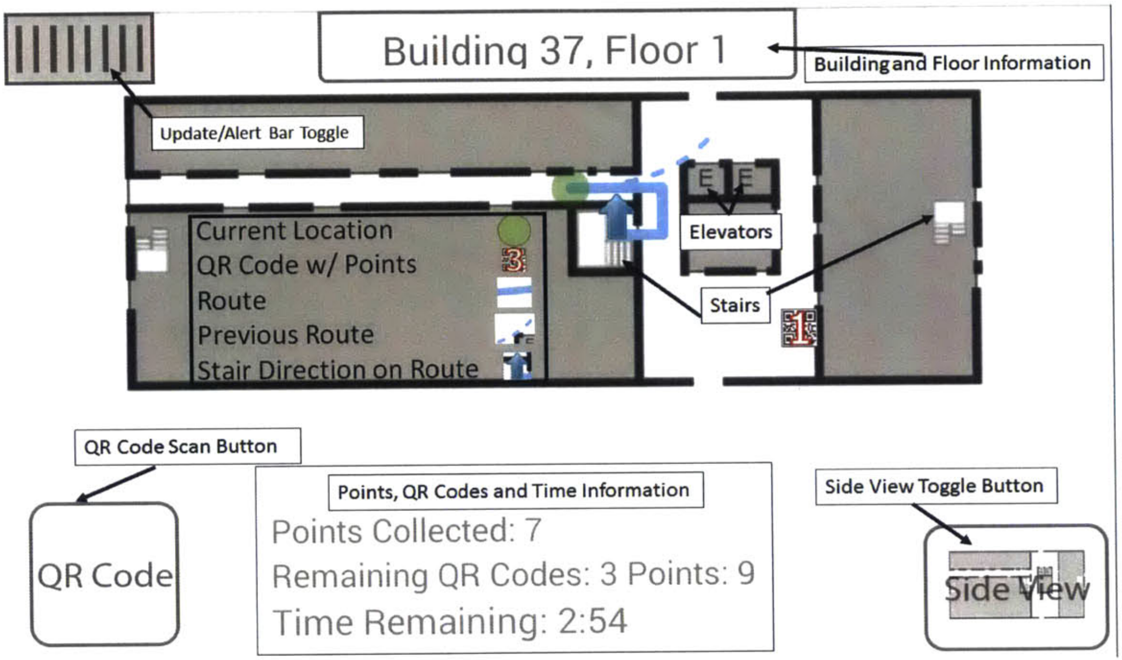


Figure 2-8: Labeled Top View Map

showed the currently-planned route. A dotted blue path showed the route that the user had already taken in this trial. The nearby QR Codes are shown with a small QR square and a red number equal to the associated point value.

Important information, to include points collected, remaining QR codes on the route and remaining time, was organized on the the Top View display below the floor plan. The Points Collected showed the number of points successfully collected during this trial. The Remaining QR Codes and Points below that, showed the number of QR codes on the route that the user had yet to scan and the total number of points they were worth if the user collected them all. There were additional QR codes with point values that were not on the route and not included in this sum. Time Remaining was a countdown clock from 180 seconds from when the trial began, following the pressing of either the choose route button or begin button.

From the Top View, the user had access to the QR Code button, Alert Bar and the Side View button. The Side View button could be pressed anytime during the trial to toggle between the Top View and Side View displays.

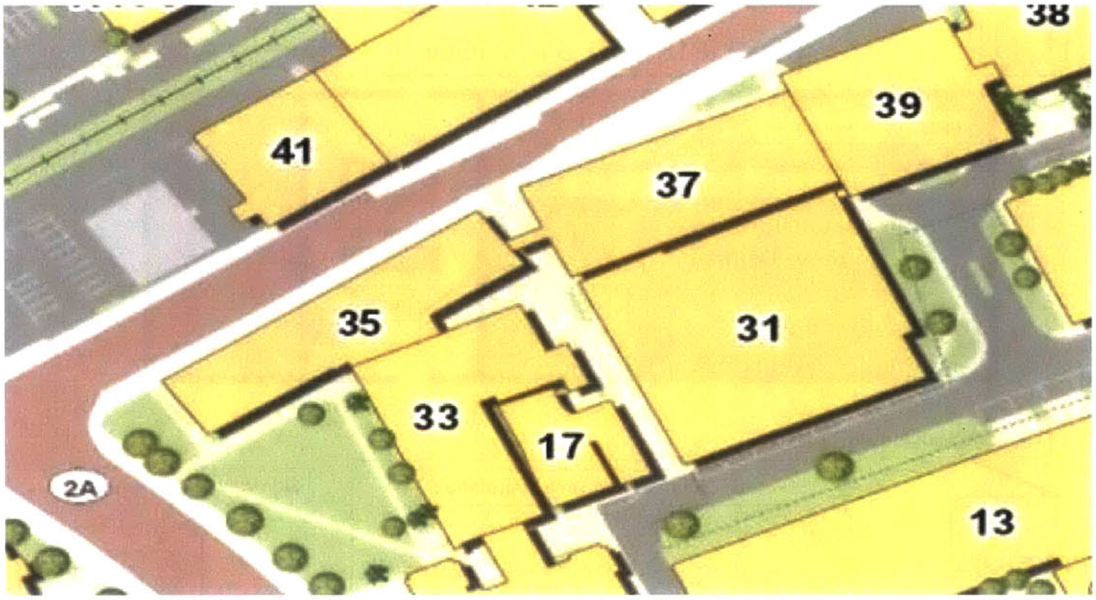


Figure 2-9: MIT map of Buildings 35 and 37

2.6.2 Side View

The Side View showed a global perspective of the trial by drawing the navigation route across the two buildings and by showing all of the available QR Code locations and associated points for the trial. From the Side View, the user could quickly reorient themselves within the trial to see, for example, how many floor changes are on the route ahead of them as they go up the stairs.

Figure 2-9 shows the actual positions of Buildings 35 and 37 relative to each other at MIT. The view provided to the user in side view was with these two buildings oriented so that Building 35 was on the left and Building 37 was on the right.

The user's current location in the building was shown with the green circle. The trial start and goal locations were shown with two stars, the start being a black star and the goal being a gold star. See Figure 2-10 for a clear view of this information with the legend in the upper left corner. The legend is not displayed in the implementation of the app as shown in Figure 2-7.

A solid blue path was drawn to show the current route, both where the user had been so far and where they still needed to go in order to reach the goal. This allowed the user to see where their path might come close to additional valuable QR Codes

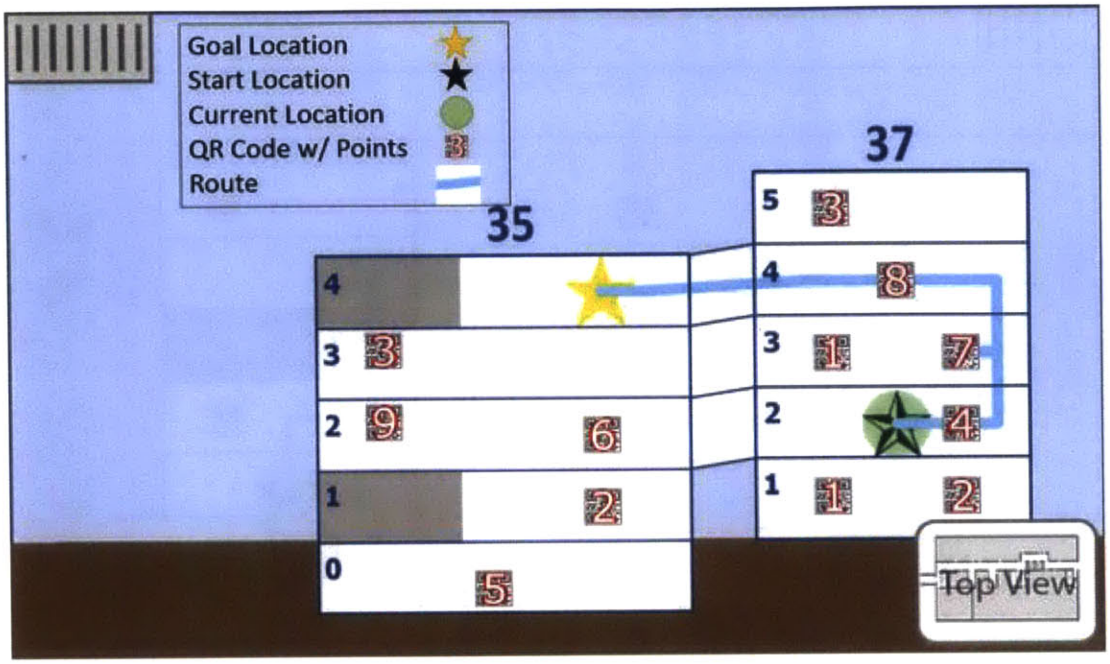


Figure 2-10: Side View with Legend

and have the ability to deviate if they had enough time to spare.

The Side View is static like the Top View in that the map does not rotate to align North nor to align with the phone orientation. The Side View button can be pressed anytime during the trial to toggle between the Top View and Side View displays. As can be seen in Figure 2-10, the user was required to toggle back to the Top View in order to scan posted QR codes. The user simply pressed the Top View button, in order to return to the Top View display where they had access to the QR Code button and important information about the current trial.

2.7 Additional Features

As previously referenced, the app had two features that were utilized in both modes. The first was the alert bar that notified the user of closed floors and provided the user a way to change their route in user-choice. The second feature was the QR code scanner which was used in every trial.

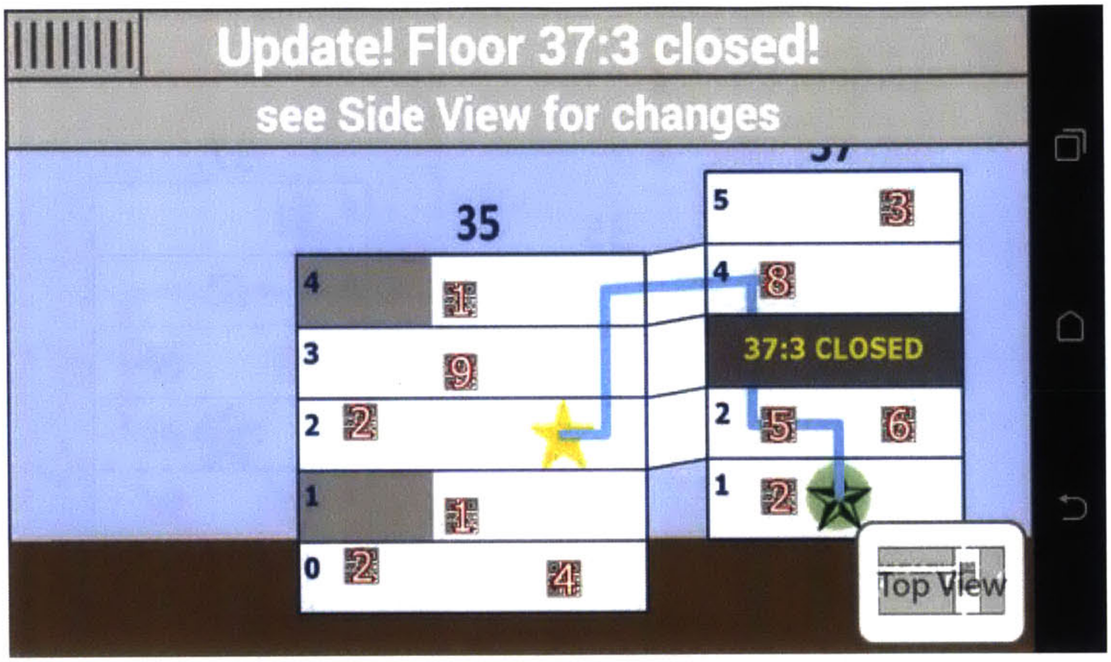


Figure 2-11: Alert Bar Example

2.7.1 Alert Bar

The Alert Bar, as shown in Figure 2-11, informed the user when an update occurred and which floor was closed. When an update occurred, the Alert Bar immediately pops up with the alert message and an alert sound is issued from the app. There were three sounds selected for this experiment: a red alert sound (notify changing routes), an industrial alarm (notify the occurrence of a floor closure) and the turn on sound of a gameboy (notify user that a route with more points has been found). The Alert Bar can be toggled by pressing the button in the top left corner. The Alert Bar also allows the user to change the path in the user-choice mode either due to the update, or just in general.

2.7.2 QR Code Scan

The QR Code button was pressed during the trial when the user wanted to scan a new QR code posted on the wall that was worth points or the end location. Not all posted QR codes had value for every trial though, so scanning every QR code would not necessarily provide the user with points. It would however, update the position

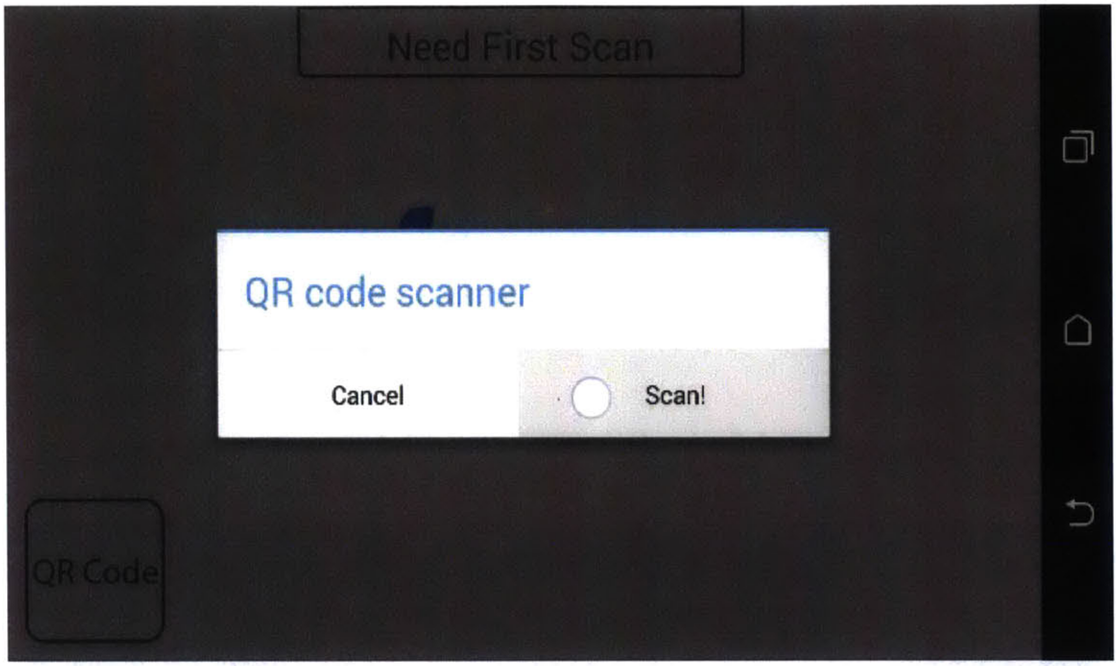


Figure 2-12: QR Code Scan View

of the user on the map regardless of point value.

To scan the QR code, the user first selected the QR Code button in the bottom left of the Top View screen, then by pressing "Scan!" as shown in Figure 2-12, the user was brought to the Barcode Scanner app. This app allowed the user to scan QR codes, by aiming the camera at the QR code posted on the wall as shown in Figure 2-13. After successfully scanning the QR code, the user was returned to the Top View.

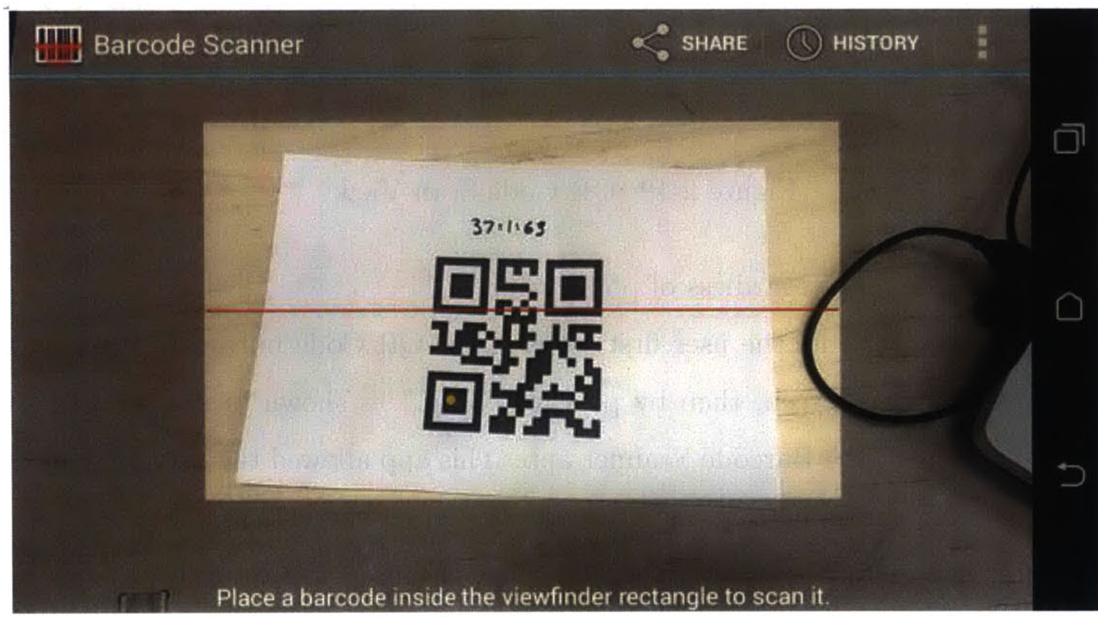


Figure 2-13: QR Code Scanning

Chapter 3

Experimental Methods

This chapter will present the experimental design, experimental measures and plan for statistical analysis. This is building on the background literature and the app design detailed in Chapters 1 and 2.

3.1 Experimental Design

This section will look primarily at the experimental design. This section is subdivided into background, participant summary, training and experimental procedures.

3.1.1 Background

The overall purpose of the experiment was to test the interactions between the user-choice and computer-select modes with and without an update within subjects and between subjects. The main effects and interactions were evaluated for quantitative performance, subjective performance, trust, situation awareness and workload metrics. Table 3.1 shows the test matrix used for this experiment.

Table 3.1: Test Matrix

User-choice with an update	Computer-select with an update
User-choice without an update	Computer-select without an update

Based on the test matrix, three main questions were examined:

1. Is there a difference in performance between computer-select and user-choice?

2. Is there a difference in performance with and without an update?

3. Is there an interaction between automation mode and update case for performance?

The expectation was that there would be significant differences for each question. The experimental measures and specific hypotheses for these questions will be broken down and explored in more depth in section 3.2.

The design of this study was a within subject design, where each subject performed each condition (as seen in Table 3.1) six times for a total of 24 trials. This allowed for a smaller number of participants as each subject could be compared to themselves.

3.1.2 Participant Summary

To be included in the study, subjects were required to meet the following inclusion requirements:

1. Between the ages of 18-34

2. Vision correctable to 20/20

3. Currently serving in a military training program, active duty or reserves

4. Passed their last physical readiness test or service equivalent

5. Must be male

These ages were selected based on the fitness standards set forth by the US Navy. The goal was to remove fitness as a confounding variable by having a set baseline. This is why males and military personnel were the targeted subject population. In addition, having all military personnel would be provide individuals from similar backgrounds and possibly mitigate user mentality as a confounding variable. Vision

Table 3.2: Subject Background Information

Subject Background Information		
N = 8 subjects	All Male	All 20/20 Correctable Vision
Preferred Time of Day:	Morning: 2	Evening: 6
	Mean	Standard Deviation
Age	21.75 years	1.753 years
Military Time in Service	4 years	1.871 years
Workout Frequency	5.625 times/week	3.114 times/week
Workout Length	63.75 min/workout	19.23 min/workout
PTA Test (> 20 is above average)	23.8	2.509
Sleep before training	7.625 hours	0.916 hours
Sleep before Experiment Trials	7.469 hours	1.322 hours

was a requirement since the sight of objects in the hallways could impact subject performance.

The experiment was completed by eight male subjects between the ages of 19-24 that were currently active duty or completing a reserve officer training program within the Boston/Cambridge MA area with 7 current MIT students and 1 Tufts student. All subjects had passed their last physical and physical readiness test conducted by their service. In addition subjects worked out a minimum of twice a week for 30 minutes. Subjects averaged 7.5 $\pm$ 1.18 hours on nights prior to training and test sections. Every subject scored above average on the perspective ability test[25] (PTA) and had vision correctable to 20/20 as reported and tested prior to the experiment. Table 3.2 provides an overview of the sample population that conducted the experiment. Subjects were recruited through the ROTC and military email lists at MIT by the recruitment notice seen in Appendix D. The study was approved by the Committee on use of Human's as Experimental Subjects (COUHES) at MIT. Each subject was required to review and sign the consent form before any part of the experiment was conducted. All eight subjects approved the consent form and completed the full study. The consent form can be seen in Appendix G.

3.1.3 Test Matrix Implementation

As described in Chapter 2, each trial was designed to be comparable to the other trials. Each of the six trials was replicated for the four conditions in the Test Matrix to form blocks. Each block was defined by one automation type and an even spread of trials with and without updates. To minimize order effects on trial, we had two different conditions. Condition 1 saw the subject experience computer-select first and last with user-choice second and third and consisted of all of the odd numbered subjects. Condition 2 had the exact opposite setup and was comprised by the even numbered subjects. The trials were administered in a random order through the first two blocks and this order was reversed for the second day, to see if there were order effects. For Condition 2, the user sees the trials in the reverse order of Condition 1 for each block. For example, Condition 1 sees the computer-select mode in the order of Lima, Golf, Bravo, Papa, India and then Charlie on day 1 and Condition 2 sees them in the inverse order of Charlie, India, Papa, Bravo, Golf and Lima in their second block of the day. The two conditions can be seen in Appendix O. Going forward, Condition 1 will be referred to as the Odd Order and Condition 2 will be called the Even Order.

3.1.4 Procedures

The experiment was conducted over a period of three days, with the first day being the training of the system and the second and third days being the experimental testing. No two sessions were completed on the same day. The maximum gap was four days between training and the first testing day and three days between testing days. The goal was to schedule them in successive days, but the subject and experimenter availability did not allow for this. The interactions of the experimenter were scripted to ensure that the same level of training and instruction was provided to each subject. That script can be seen in Appendix F. Prior to the study, each subject was evaluated for the inclusion criteria via an email conversation. If the inclusion criteria were met, then they were scheduled for a training day and two testing days. The screening criteria can be seen in Appendix E and were listed in Section 3.1.2.

Training Day

The purpose of the training day was to collect basic information about the sample population and introduce the subject to the app and experiment to a sufficient level to reduce learning effects. Participants completed a pre-training and post-training survey (see Appendix I), as well as a vision test and the Perspective Taking Ability Test[25] with the instructions in Appendix L. All surveys were performed before the training to mitigate boredom except for the post-training survey.

The subject then read through the training document (see Appendix M), NASA TLX instructions (see Appendix K)[13], types of survey questions document (see Appendix H), and the SA items sheet (see Appendix N, Figure N-1) for review. In addition the three noises used in the app were simulated. This was also the time where the subject would report the amount of sleep they had logged the previous night. The questions on the last page of the training document were used to determine the comprehension of the task.

The subject was then given a 20 minute active walk-through buildings 35 and 37 to get a feel for the locations of the QR codes and the stair access points using trail call sign Quebec. When the subject felt sufficiently comfortable with the buildings, four practice trials were completed (call signs Mike/November) that spanned each condition of the study. With all trials, an experimenter followed at a slight distance in case any issue arose.

Subjects were considered fully trained following the completion of the practice trials and the post-training survey. The review questions from the training document were asked again to ensure retention and a final opportunity was provided for questions. The training day was about 1.5 hours in duration.

Experimental Test Day

The experimental testing days began by reviewing the overall task and objectives, while allowing the participant to ask questions. The three priorities were:

1. Scan the goal QR code within 180 seconds of beginning the trial

2. Maximize the amount of points they collect

3. Accomplish the secondary SA task

Trials were not repeated if something unexpected occurred like a fire alarm or an app failure. In the case of this experiment there were 12 trials removed for app failure. The distribution of the failures can be seen in Figure 3.3.

Each subject had 180 seconds for each trial to navigate through the buildings, scan QR codes for points, and scan the final goal QR code. Each trial consisted of the subject being directed to the starting node and handed the app with the correctly load scenario from the experimenter. The subject would then start the trial whenever they were ready. Following the conclusion of the trial, subjects were given surveys corresponding to the trial they were on as shown in Appendix I. The subject would first complete the electronic survey on the app and then the written SA survey at the goal location. Following every third trial, the subject would then be directed back to the experimenter's office for the completion of the NASA TLX. Once all of the surveys were concluded, the subject would be directed to the next start location until all trials were completed for the day.

Overall, each trial from walking to the start, planning, completing the trial and completed subsequent surveys took about seven minutes. Following every third trial the subject had an additional ten minutes of rest while the SA characters were switched out. The full protocol for one test day was approximately two hours.

All survey questions were digitally administered and stored except the written SA questionnaire. In addition, live trial data was collected via the app for the time to reach the end goal, plan time, top view time, side view time, QR scanner time, points, penalties, alerts, route changes, QR codes scanned, the route traveled and system interactions.

Table 3.3: Distribution of App Failures

Distribution of 12 App Failures	
	Total Failures
Day 1	7
Day 2	5
No Update	4
Update	8
Computer-select	0
User-choice	12
Odd Ordered	2
Even Ordered	10
Trial 1	2
Trial 2	2
Trial 3	3
Trial 4	4
Trial 5	0
Trial 6	1

3.1.5 Completing a Trial

As previously mentioned, subjects had 180 seconds to navigate through the buildings, scan QR codes for points, and scan the goal QR code. Their first priority was reaching the goal QR code and scanning it before the trial time is up. Their second priority was maximizing points by scanning additional QR codes along your route. Their third priority was the secondary situation awareness task which will be explained in Chapter 3. This section will discuss the actual use of the app in computer-select and user-choice modes.

Computer-Select Navigation Trial

In the computer-selected mode, the app performed all route planning functions to include rerouting due to updates and deviations from the path. The user was informed of all route changes and updates by the app.

To begin the trial, the experimenter scanned the start QR code to set the start location. The user than had time to plan out their route. The trial time began when the user pressed Begin from the initial Side View.

The first map the user saw after selecting Begin was the Top View. From that point on the trial was live and the subject had full use of the app to meet his objectives within the time limit.

If a new navigation route was calculated for them, a sound notification played and a message appeared in the Alert Bar at the top of the display. The new route did not necessarily include the same floor changes and QR codes so it was recommended that they toggle to the Side View to view the new route.

User-Choice Navigation Trial

In the user-choice mode the user selected the route that they wanted to use for the trial. The user-choice mode would not change the navigation route without the user manually selecting a new option. If he wandered too far off of their selected route, then they were prompted to continue on their current route or update to a new shorter

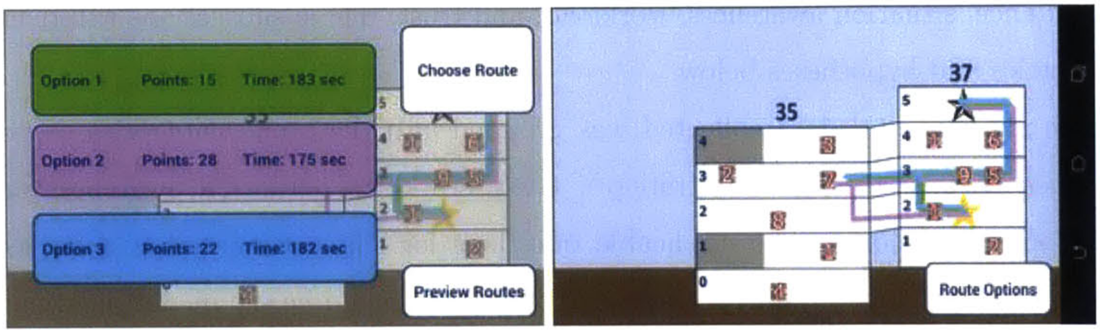


Figure 3-1: Side View with User-choice route options

route.

To begin the trial, the experimenter scanned the start QR code to set the start location. The user than had time to plan out their route. The trial time did not begin until the user choose a first route to use for navigation.

The user was given three different route options to choose from to start the trial, shown in the User-choice Side View. Each route option would pick up a certain number of QR Code points and also have a time estimate associated with it. Figure 3-1 demonstrates the views provided to the user to select their route.

The Side View was shown behind the route options and could be viewed by pressing the Preview Routes button in the lower right hand of the screen. Each route was color coded with the color of the respective route option buttons. In the above figure for example, route option 1 has a green button and is shown with a green path on the Side View.

The user selected which route they wanted to use in the trial by clicking on it and then confirming the choice by clicking on the Choose Route button. The trial time began as soon as the Choose Route button was pressed. As with computer-selected mode, after the user selected Choose Route, the first map seen was the Top View.

3.2 Experimental Measures

This section will look primarily at the experimental measures used to test the experimental hypotheses. This section is subdivided into overall performance, subjective performance, situation awareness, workload, and trust. These subsections will define the metrics and hypotheses below.

The experimental data collected was grouped into the broad categories of performance (scores, timing and strategy), trust, workload, situation awareness and background questions. The applicable questions for these categories can be seen in Appendix J. The overall hypothesis was that there would be significant differences between main effects of mode, of updates, and the interaction between mode and update.

Background information collected to describe the sample population included sleep, GPS knowledge, PTA, training effectiveness, eye sight, fitness, morning or evening person and military background. These data were used to mitigate confounding factors between the subjects.

The ROTC and military population was selected to minimize variance of fitness levels for endurance and speed in the timed navigation task, since they must regularly pass a basic fitness test. In addition subjects would be conducting the experiment through a similar mindset and shared experience. The PTA test was conducted to determine the subjects mental capability to orient themselves, since the app's map was static in orientation. The six unique trials with the inclusion of updates and the time sensitive nature of the task, created an environment of stress and dynamic complexity. Having each subject complete every trial under the same conditions, allowed for the complexity and stress to be held constant. This complexity and stress through pressure from temporal demands and point maximization, was chosen to make the experiment have more realism. Training was consistent for each subject as described above. Subject learning was was assessed by comparing the two experimental days and by the subjects reporting any strategy changes. Vision variance was controlled by requiring a correctable vision of 20/20, to ensure that observations of characters and collection of QR codes was not impacted by vision variations.

The overall hypotheses for this study were:

1. H1: There is a difference in user performance between adaptive (computer-select) and adaptable (user-choice) automation

2. H2: There is a difference in performance with and without an update to the scenario

3. H3: There is an interaction in performance between automation type and update case

3.2.1 Measures of Objective Performance

Objective performance was determined by the live data collected by the app which could be distributed into three main groups: overall scores (points collected and penalties), time in app modes (top, side and QR scanner views, planning and arrival) and task strategy (alerts, route changes, system interactions and strategy changes).

Penalties were given if the subject was unable to scan the goal QR code before the trial time had run out. The subject lost a fourth of their points for being up to 10 seconds late. After that, the penalty became half of their points. Anything more than 30 seconds late was considered a failed trial and the subject received no points. There was no penalty or bonus for arriving early.

There were also penalties for using closed-off areas within the two buildings, including floors that might become unavailable after receiving an update. A deduction of 5 points was given for crossing a closed floor. There was no deduction for subjects that were already on the removed floor when the update occurred as long as the subject left immediately. The staircases of a closed floor were still permitted. In addition elevators were not permitted during any trial and a deduction of all points would be levied.

For the subject's safety as well as that of faculty and students who may be in the hallways during the experiment, the subject was not to exceed the speed of a fast walk at any time. They would be assessed a penalty of half of their points for exceeding on average 4.5 mph which is approximately the preferred transition from walking to running[40].

All penalties were assessed to subjects for failure to scan the goal QR code before time expired. Penalties were applied following the trial. A summary of all possible

Table 3.4: Penalties

Violation	**Penalty**
Late <10 seconds	**¼ Point Total**
Late 10 - 30 seconds	**½ Point Total**
Late >30 seconds	**All Points**
Cross Closed Off Area	**5 Points**
Average > 4.5 mph	**½ Point Total**
Use an elevator	**All Points**

penalties can be seen in Table 3.4.

Overall scores were analyzed by looking at five metrics: total points collected before and after penalties, delta first route points before and after penalties and penalties. Delta first route points looked at the difference between the total points collected and the projected point total of the subject based on the initial route selected. For example, Subject A selects a route in user-choice worth 28 points, but only collects 23 points during the trial. Thus his delta first route points would be -5.

It was hypothesized that computer-select mode would enable the user to collect significantly more points than user-choice mode with and without an update through each of the four point metrics. In addition subjects were expected to be able to collect significantly more points in trials without update, regardless of mode. The hypothesis was that the difference between the two modes would be greater in trials with an update.

Time analysis was based on six metrics: top view time, side view time, QR code scanner time, planning time arrival time and ratio between top view and arrival time. Each time was logged during the trial. Planning time consisted of the time between the experimenter handing the prepared scenario to the user and the subject pressing the begin button. Top view, side view and QR code scanner time correspond to the amount of time that the user was in that view. Finally arrival time was the time logged by the scanning of the goal QR code. This time is the summation of the time spent in top view and side view. The QR code scanner time was outside of the overall trial time and actually paused the trial clock. This was not revealed to the subjects and at no time was it apparent to the experimenter that they were aware of this bonus time. Since top view and side view encompassed the arrival time, a ratio of top view time and arrival time was analyzed as a way to normalize across subjects and trials.

There were two primary hypotheses associated with timing. There was expected to be significantly more time spent in side view than top view in the user-choice mode and when an update occurred. The other time metrics collected were analyzed for effects post-hoc.

Objective strategy analysis was based on four metrics: alerts, route changes, system interactions and strategy changes. Alerts was intended to track the amount of route change recommendations by the app, but the logging of these was over-reported. For example the log file recorded five values in under a second for a majority of the trials, thus there alert values in the single digits and hundreds for a 180 second trial. Reports at that speed would not allow for subject response and were thus discredited. Route changes were the quantity of changes the app made during the trial. In computer-select, this was logged for every change, while in user-choice, this was logged whenever the user agreed to change their route or there was an update. System interactions was the log of all buttons pressed throughout a trial. Finally, the direct part of strategy changes was the binary reporting of a change by subjects.

It was hypothesized that there would be more route changes and alerts in computer-select mode than user-choice mode regardless of update and when an update occurred regardless of mode. It was also expected that the number of system interactions would be greater for trials with updates and in the user-choice mode. Finally it was thought that subjects would have more strategy changes following updates.

3.2.2 Measures of Subjective Performance

Subjective performance was determined by six metrics: mode preference, best/worst part of the app, strategy changes, self-evaluation of performance, task realism and task complexity.

Mode preference, best/worst part of the app and strategy changes were open ended questions that allowed for any subject response. Self-evaluation of performance, task realism and task complexity were Likert items that allowed the subject to rate them from low to high. The lone hypothesis was that the participants would have a preference toward user-choice mode.

3.2.3 Measures of Situation Awareness

Situation awareness questions were used to determine how much knowledge the subject had of their environment and the scenario. Following each trial, subjects were asked about their memory of the trial just completed to include: unclaimed points available during the trial, the total of unclaimed points in the trial, the initial route, floors visited, and the start and goal locations. In each hallway there were placed one of 13 different characters which can be seen in Appendix N. In every trial subjects were asked about the locations of any characters they may have seen in the halls. Each subject had a handout of all 13 characters while completing the survey. The characters displayed in the halls, their locations during the study, the written survey and an example of a completed survey with grading can be found in Appendix N. The characters were switched every third trial. The written survey was also used to mark floors the subjects thought they visited and their start and end locations. Of the questions asked five were related to the SA level of perception (start, goal, floors visited, initial points on their route and if they saw any characters) and three related to the SA level of comprehension (which characters, where, and the total number of unclaimed points).

These SA questions required subjects to report back information following a trial, which made the task more realistic and comparable to a scout mission. This task can be heavily dependent on memory, but the alternative of interrupting the trial to acknowledge a message or make the subject stop every time a character was passed would negatively impact the primary task. It also shows how much awareness the subject had beyond the device. In addition, the task forces the subjects to be on the look out for items not marked on their map, which would hopefully mitigate tunnel-vision between each QR code and reduce the subjects pace. By reducing the subject's pace, the task would also help mitigate the disparity in fatigue between the first and last trials.

Situation awareness consisted of seven metrics: start location, goal location, characters seen (any, total, correct, correct floor and correct location), floors with scanned QR codes, delta initial points between reported and actual, points unclaimed and composite SA.

The written SA survey comprised the information for start location, goal location, characters seen (any, total, correct, correct floor and correct location) and floors with scanned QR codes. An example of SA scoring of the written sheets can be seen in Appendix N. Delta initial points was the difference between the number of points the subject recalled were on their initial route and the number of points actually on their initial route. Points unclaimed was intended to look at the global amount of points still remaining in the scenario when the subject completed their trial, but the framing of the question led to inconsistent responses with some subjects looking at the global map, while others looked at their initial route. Therefore this question was removed from analysis. Finally composite SA was the sum of the binary totals of start, goal and seeing any characters and the total amount of characters seen, characters correctly seen, characters placed on the correct floor, characters placed in the correct location, correctly reported floors minus unreported floors with a scanned QR code.

For situation awareness, it was hypothesized that subjects would do well with marking the start and end location and floors visited, but the observance of characters would be minimal relative to the number of floors the subject visited. There was hypothesized to be an improvement on the second testing day. It was also hypothesized that subjects would have better performance on the SA task during trials without an update and in the computer-select mode.

3.2.4 Measures of Workload

Workload was used to track the difficulty of the task and the amount of fatigue experienced throughout the 12 trails in a given day. Workload was comprised of six metrics: NASA TLX (composite, weighted and raw), physical and mental demand, physical and mental assistance from the app, and fatigue.

The NASA TLX[13] is a standard workload test that determines a composite workload score based on raw scores for mental, physical, temporal, performance, effort and frustration components that are weighted based on subject's pairwise comparisons. Physical and mental demand, physical and mental assistance from the app, and fatigue were Likert items that allowed the subject to rate them from low to high.

It was hypothesized that workload would be high initially due to re-acquaintance with system at the beginning of experiment days and towards the end of the day due to fatigue increases. In addition, it was hypothesized that workload would be higher in user-choice mode and in trials with an update.

3.2.5 Measures of Trust

Trust was used to determine the subject's use of the app in both modes. Trust was comprised of nine metrics: system trust, trust of route changes, expectation of route changes, clarity of the app's actions, trust of the refresh rate, understanding of the reason for the route change, ease of identifying the change, usefulness of the app and a trust composite.

The following were scored using Likert items: system trust, trust of route changes, expectation of route changes, clarity of the app's actions, trust of the refresh rate, understanding of the reason for the route change, ease of identifying the change, and usefulness of the app. The subject rated the Likert items from either low to high or strongly disagree to strongly agree. The trust composite was the sum of these eight Likert items to create a Likert scale.

It was hypothesized that there would be a difference in trust for the two automation types, but the experimenters were unsure which way it would favor. It was also hypothesized that there would be greater trust at the end of the study than at the beginning.

3.3 Summary of Hypotheses

Here is a summary of the specific hypotheses that were made prior to the study. They will be examined in detail in Chapter 4.

1. Performance

   (a) Overall Scores

      i. PH1: The computer-selected automation condition will collect significantly more points without an update than with an update

      ii. PH2: The user-choice automation condition will collect significantly more points without an update than with an update

      iii. PH3: The computer-selected automation condition will collect significantly more points with an update than user-choice automation

      iv. PH4: The computer-selected automation condition will collect significantly more points without an update than user-choice automation

   (b) Timing

      i. STH1: There will be significantly more time spent in side view than top view in user-choice automation trials than computer-selected automation trials

      ii. STH2: There will be significantly more time spent in side view than top view in trials with updates than trials without updates

   (c) Strategy

      i. SRH1: There will significantly more route changes when using the computer-selected automation condition with and without updates than user-choice automation

      ii. SRH2: There will significantly more route changes when an update occurs than when no update occurs

      iii. SSH1: There will be a significant increase in system interactions in trials that include an update for both automation types

iv. SSH2: There will be a significant increase in system interactions in trials that include user-choice automation than computer-selected automation

v. SCH1: There will be significantly more strategy changes following updates than a trial without an update

2. Situation Awareness

   (a) SAH1: There will be significantly better performance on the secondary task on trials without an update

   (b) SAH2: There will be significantly better performance on the secondary task on trials in the computer-select mode

3. Workload

   (a) WH1: There will be a significantly higher workload with user-choice automation for the NASA TLX with respect to computer-selected automation

   (b) WH2: There will be a significantly higher workload with computer-selected automation from the subjective tests with respect to user-choice automation

   (c) WH3: There will be a significantly higher workload during the trials with an update

   (d) WH4: Fatigue will significantly increase throughout the study and will negatively impact the performance metrics of time and points collected while increasing number of route changes

4. Trust

   (a) TH1: There will be a significant difference in trust of the two automation types

   (b) TH2: There will be a significant difference in trust at the beginning as compared to the end of the experiment

5. System Preference

   (a) MH1: There will be a user preference for user-choice automation

3.4 Statistical Analysis

This section will look primarily at the statistical analysis plan implemented on the collected data. This analysis will be presented in Chapter 4.

Statistical analysis was completed using Systat statistical analysis software. Mixed linear regression models were fitted to examine effects for mode, update, day, order and trial; as well as to include possible interaction effects. The linear regression models were validated using the Kolmogorov-Smirnov (KS) test for normality and Levene's test for constant variance for the model's residuals. Counting variables and Likert items were analyzed using a non-parametric Pearson Chi-squared test. Kruskal-Wallis and the Mann-Whitney U tests were applied to determine if there was a difference between two or more groups of non-parametric data if the variable was not easily grouped to allow for the Pearson Chi-squared test. The time variables of top view, side view, planning and QR code scanner were transformed by natural log to allow for analysis because the time data followed a lognormal distribution.

Chapter 4

Results and Discussion

This chapter will present the results of the statistical analysis, as outlined in Chapter 3 and discuss the implications within the scope of current literature. This chapter will first look at the general description of the sample population and then describe performance, SA, workload, trust and preference in that order.

4.1 Independent Variables and Regression Models

As shown previously, Table 4.1 details the primary independent categorical variables that were included to assess main effects, learning, as well as those associated with order. The coding value shows the order viewed within the mixed regression for categorical variables. Within the regression tables, the estimates, $\hat{\beta}$, represent the coefficient for that variable with that variable representing a 1 or -1 based on the coding with the first term within the binary variables reflecting the 1 and the other term representing -1. For example, day 1 was always coded as a 1 and therefore day 2 would be coded as a -1. Trial, which had six terms, was represented by five variables (x1 + x2 + x3 + x4 + x5) with each term being utilized as a 1 for its own trial or otherwise a 0 except for trial 6 which was coded as a -1 for each of the five trial terms. For example, an estimate for trial 1 would include x1, while all other terms would have zero effect. Graphical examples for mixed regression model interpretations can be found in Appendix Q. Figures showing the significance for investigated main and

Table 4.1: Independent Variables

	Groups (Coding)	Effects
Mode	Computer-select (0) & User-choice (1)	Main
Update	With (1) or Without (0)	Main
Day	1 or 2 (Day number)	Learning
Mode and Trial Order	Odd (1) or Even (0)	Order
Trial	1 to 6 (Trial number)	Learning

Table 4.2: Subject GPS Use

	Average (Between 1 & 5)	SD
1. I am confident in the task presented to me.	4.550	0.513
2. The task was clearly explained.	4.700	0.360
3. The task is applicable to a real scenario.	3.820	0.838
4. GPS route changes are appropriate.	3.650	1.043
5. GPS's actions are clear to me.	4.080	0.585
6. Changes in the route displayed on the map were expected.	3.060	0.730
7. GPS provides up to date information	3.720	0.924
8. The reason for route changes is clear.	4.000	0.878
9. GPS updates in the route are easily identifiable.	3.670	1.008
10. GPS is useful for navigating.	4.880	0.181

interaction effects from the regression can also be seen in Appendix P.

Subject's trust baseline of navigation devices was determined by asking GPS specific questions that were related to those asked throughout the experiment but in reference to the app. These were correlated to see if there was any effect of natural subject trust on the reported trust in the app. These will be examined in section 4.4 below. Following training, subjects reported sufficient confidence in the task required of them to proceed. In addition, subjects felt that the experiment generally represented a possible real scenario as shown by number 3. These results are displayed in Table 4.2, where 1 signified strongly disagree and 5 signified strongly agree.

4.2 Performance

This section will look primarily at the subject's performance in the primary task of arriving on time, while collecting the maximum amount of points. This section is subdivided into overall scores, timing, and strategy metrics.

4.2.1 Overall Scores

Before Penalties

Total points collected before penalties was applied to a mixed linear regression model (Table 4.3). There was an improvement of 1.42 on day 2, a decrement of 2.12 points when the trial had an update, and a significantly greater improvement in user-choice between days than computer-select. The interaction effect estimates that computer-select on day 1 and user-choice on Day 2 collected 1.47 additional points, while computer-select on day 2 and user-choice on day 1 collected 1.47 points less. For example, the model would estimate that a user in user-choice on day 2 without an update would collect 5.01 above the mean. Whereas, a user in computer-select on day 2 without an update would only collect 2.07 points above the mean. This means that the model expects a user to collect 2.94 more points on day 2 for user-choice as compared to computer-select for the same update case. For day 1, computer-select is expected to outperform user-choice by 2.94 points. This was the result of improvement in the use of user-choice because computer-select scores stayed relatively constant between the two days. Finally Trial 3 was shown to be a significant indicator of point total within the model due to the small variance. The Trial 3 map scenario, as shown in Appendix C, provided less natural route options to the user which resulted in the singular path that most subjects followed.

Overall these results support the idea that there was learning for user-choice between the two experimental days as the subjects improved their scores. This improvement implies that user-choice was more complex to understand and use effectively. While the computer-select operation was fairly consistent between days. In addition, this supports the premise that the update in the middle of the trial significantly

Table 4.3: Total Points Collected Before Penalties

Variable	$\hat{\beta}$	SE	Z	p
Intercept	24.04	1.75	13.74	<0.001
Day	-1.42	0.34	-4.15	<0.001
Update	2.12	0.34	6.23	<0.001
Trial 1	-0.69	0.76	-0.91	0.363
Trial 2	-0.73	0.76	-0.96	0.336
Trial 3	3.76	0.77	4.86	<0.001
Trial 4	-0.01	0.78	-0.12	0.990
Trial 5	-1.45	0.74	-1.96	0.051
Mode*Day	1.47	0.34	4.30	<0.001

affected the subjects ability to maximize their point total.

The delta between a subject's points on their first route and the amount of points actually collected was analyzed using a mixed linear regression model (Table 4.4). There was an improvement of 1.54 on day 2, a decrement of 2.18 points when the trial had an update, an improvement of 3.42 in computer-select and an order effect with even subjects performing 2.13 above the mean of 2.1 points greater than the subject's initial route. There were also two interactions and a covariate of plan time. The first interaction effect estimates that computer-select on day 1 and user-choice on Day 2 collected 1.51 additional points, while computer-select on day 2 and user-choice on day 1 collected 1.51 points less. For this model, there was an additional interaction effect between mode-update-day of 0.90 points. This effect is positive if the coded values multiple to 1 and negative if they multiple to -1. There is a graphical representation in Appendix Q. The covariate effects the predicted outcome by 0.06 improvement for each second of planning.

Overall this tells us that users were more ambitious in their route selection when given the choice, even though there was not a significant difference in the means between modes for total points collected overall, because the differences between day 1 and day 2 leveled out. As will be shown after penalties are applied, this effect is misleading. The impact of route selection is impacted by the optional routes provided. Of the three routes provided to the user, the additional routes were more ambitious than the computer-select route. Thus subjects were more inclined to select these routes. The inclusion of an update negatively impacted subject's ability to out pace their initial route. Subjects learned how to use the system better by day 2 and thus, they surpassed their initial route's projected points better on day 2. As with total points collected, most of the gain over the two days was seen in the subject's ability to operate user-choice. When broken down to include update, subjects improved in user-choice with and without an update, as compared to day 2. Subjects actually performed worse in computer-select without an update on day 2.

There was a significant relationship between the amount of planning time utilized and the subsequent number of points the subject achieved beyond their initial route. This speaks to the subjects improved general knowledge of the scenario map by planning longer which allowed the subject to reference the app less and thus move faster through the scenario. There was also found to be an effect of order. This could be due to a couple reasons. First, it could be that subjects in the odd group had more time to acclimatize themselves to the system in computer-select mode on day 1, where they are being asked to do less and then the transition is more gradual. This way, the odd subjects are more content to follow the single computer-select path and are doing less initial deviation planning. While the even subjects are bombarded at the beginning with these possibilities and start off more ambitiously. The order effect could also just be coincidence based on random assignment to the two groups. As will be shown shortly these differences could also be accounted for by even subjects being more aggressive in attempting to arrive just in time, which resulted in more penalties and more severe time penalties as discussed below.

After Penalties

There were 19 total trials in which subjects were penalized. All penalties were related to failure to complete the trial in under 180 seconds. Table 4.5 shows all possible delineations of these trials. There was found to be only an effect of day ($X^2 = 4.993$ and $p = 0.025$) on the number of penalties using a Pearson Chi-squared test with most of the penalties occurring on day 1 as shown in Figure 4-1. This speaks to subjects adjusting to the time constraint and the system's time estimate. It could

Table 4.4: Delta First Route Points Before Penalties

Variable	$\hat{\beta}$	SE	Z	p
Intercept	2.10	1.34	1.57	0.116
Day	-1.54	0.36	-4.27	<0.001
Update	2.18	0.36	6.11	<0.001
Order	2.13	1.08	1.97	0.049
Mode	3.42	0.38	9.09	<0.001
Plan Time	0.06	0.02	3.51	0.002
Mode*Day	1.51	0.36	4.22	<0.001
Mode*Day*Update	0.90	0.36	2.52	0.012

also be related to more knowledge of the experiment environment.

When looking at total points collected after penalties are applied, the mixed linear regression model is very similar to the model before penalties (Table 4.6). There is an improvement of 1.95 on day 2, a decrement of 2.3 points when the trial had an update, and a significantly greater improvement in user-choice between days than computer-select. The interaction effect estimates that computer-select on day 1 and user-choice on Day 2 collected 1.52 additional points, while computer-select on day 2 and user-choice on day 1 collected 1.52 points less. For example, the model would estimate that a user in user-choice on day 2 without an update would collect 4.78 above the mean. Whereas, a user in computer-select on day 2 without an update would only collect 3.72 points above the mean. This means that the model expects a user to collect 1.06 more points on day 2 for user-choice as compared to computer-select for the same update case. For day 1, computer-select is expected to outperform user-choice by 5.02 points. Unlike total points, there was no effect of trial and there was the additional effect of mode. Here computer-select increased point totals of subjects by 0.99 points above the mean.

Overall, this model is more indicative of the subject's actual performance in the experiment since penalties are now included. There still remains an effect of learning experienced over the two days. As before this is mostly a byproduct of the subject's improvement in the utilization of user-choice mode on day 2. Unlike before, there is now a significant effect of mode with subjects collecting more points in computer-select. Since the only difference between the two models is the application of penalties

Table 4.5: Distribution of Penalties by Independent Variables

Distribution of Penalties by Groupings					
Penalty Amount	None	Quarter	Half	Full	Total
Day 1	75	9	2	3	14
Day 2	86	3	1	1	5
No Update	84	5	1	2	8
Update	77	7	2	2	11
Computer-select	89	6	1	0	7
User-choice	72	6	2	4	12
Odd Ordered	87	6	0	1	7
Even Ordered	74	6	3	3	12
Trial 1	27	2	1	0	3
Trial 2	27	3	0	0	3
Trial 3	25	1	1	2	4
Trial 4	26	2	0	0	2
Trial 5	28	3	0	1	4
Trial 6	28	1	1	1	3

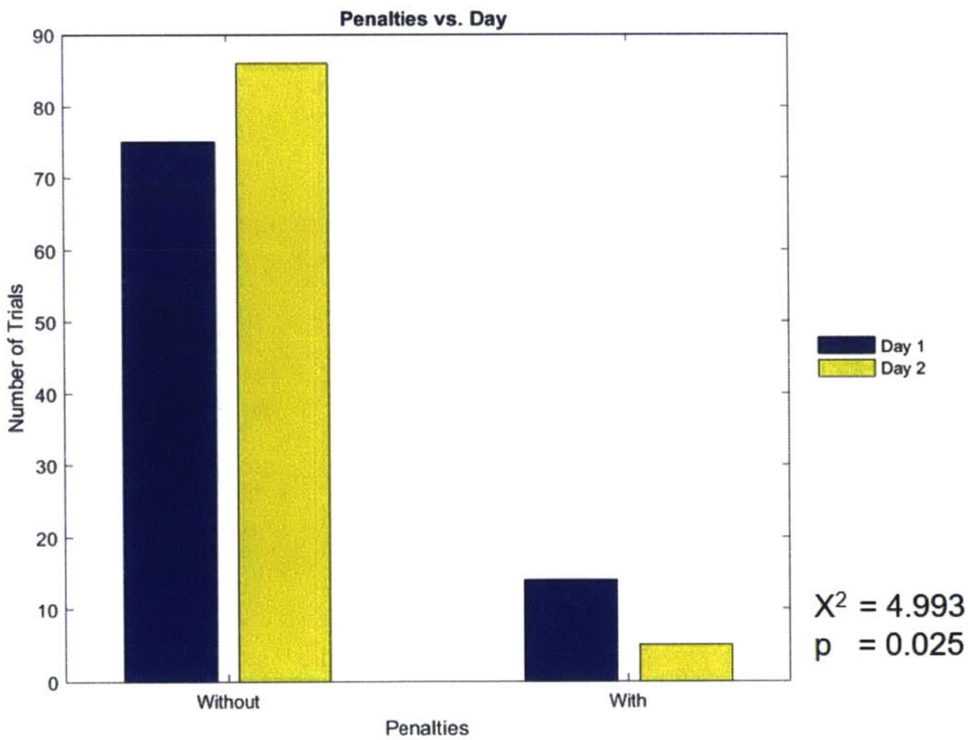


Figure 4-1: Penalty Contingency Table Results

Table 4.6: Total Points After Penalties

Variable	$\hat{\beta}$	SE	Z	p
Intercept	22.55	1.54	14.63	<0.001
Day	-1.95	0.44	-4.39	<0.001
Mode	0.99	0.45	2.23	0.025
Update	2.30	0.44	5.20	<0.001
Mode*Day	1.52	0.44	3.41	0.001

upon the results, this suggests that users were operating the two modes differently. In user-choice, the users were more aggressive and attempted to collect more points with the more ambitious routes, but accrued more penalties and thus diminishing the overall point total to the point that computer-select was estimated to provide more points. As before there is still a significant decrement in point collection when an update occurred as expected, since less points would be available to the subject. The effect of Trial 3 was removed from the model once penalties were included. This supports the design of the trials as comparable.

The delta between a subject's points on their first route and the amount of points actually collected after penalties was analyzed using a mixed linear regression model (Table 4.7). There is an improvement of 1.91 on day 2, a decrement of 2.21 points when the trial had an update, an improvement of 3.86 in computer-select above the mean of 3.18 points greater than the subject's initial route. There were also two interactions. The first interaction effect estimates that computer-select on day 1 and user-choice on Day 2 collected 1.48 additional points, while computer-select on day 2 and user-choice on day 1 collected 1.48 points less. For this model, there was an additional interaction effect between mode-update-day of 1.2 points. This effect is positive if the coded values multiple to 1 and negative if they multiple to -1. There is a graphical representation in Appendix Q.

Overall, these data support that subjects generally outperformed the computer-select route supplied to them. This model indicates that the subjects were indeed more ambitious in their route selection in user-choice mode. The occurrence of an update negatively impacted the subject's performance relative to their initial route as expected. There were still learning effects after penalties, with the subjects out pacing

Table 4.7: Delta First Route Points After Penalties

Variable	$\hat{\beta}$	SE	Z	p
Intercept	3.18	1.59	2.00	0.046
Update	2.21	0.45	4.93	<0.001
Mode	3.86	0.45	8.57	<0.001
Day	-1.91	0.45	-4.25	<0.001
Mode*Day	1.48	0.45	3.28	0.001
Mode*Day*Update	1.20	0.45	2.67	0.008

their initial route more effectively on day 2 due mostly to improvement in user-choice. As before the subjects saw a slight decrement in using computer-select without an update, but improved with an update to make the delta negligible when pooled. The improvement in user-choice was in both the update and no update cases over the two days. Unlike the before penalty model, there was no covariate of planning time and no order effect. This supports the notion that the even subjects outperformed their initial route by violating the arrival time.

Overall, these results support PH1 and PH2 and partially support PH3 and PH4. The partial support is due to the interaction effect between mode-day. The delta first route points though, help explain that this was most likely due to the better initial paths being provided. A follow on study where the user selects a baseline of initial route planning parameters and deals with the computer-select functionality would be informative.

4.2.2 Timing

Planning time was defined as the time between the user receiving the phone with the prepared scenario and the initializing of the trial. Plan time was initially looked at as a linear regression, but failed the Kolmogorov-Smirnov normality test and Levene's test for constant variance. The data for plan time was transformed using natural log and a mixed linear regression model was fitted. The natural log of plan time was found to have significant effects of mode and day with an interaction effect between day-order. These can be seen in Table 4.8. As shown in the table, users spent more time planning in user-choice mode of 0.2 percent increase and on day 1 of 0.1 percent

Table 4.8: Natural Log of Planning Time

Variable	$\hat{\beta}$	SE	Z	p
Intercept	10.31	0.26	39.63	<0.001
Mode	-0.20	0.03	-7.00	<0.001
Day	0.10	0.03	3.45	0.001
Order*Day	-0.11	0.03	-3.93	<0.001

increase. Whereas the even subjects spent similar amounts of time planning both days, an additional term of the interaction effect was required to fully characterize this. The interaction effect estimates that even on day 1 and odd on day 2 spent 0.11 percent less time planning than the mean, while even on day 2 and odd on day 1 spent 0.11 percent more time planning than the mean.

Overall, this shows that subjects spent more time analyzing the three initial routes in user-choice than the single computer-select route. In addition, this appears to demonstrate the different mentality between the odd and even groups. The even group continued to meticulously plan their route on day 2, while the odd group decided to be more decisive in their selection. As shown in the previous section, this did not lead to a significant order effect for the amount of points collected.

The amount of time user's spent in top view was also transformed using natural log, because the model failed normality and constant variance. There was found to be an effect for trials 1, 3 and 5 as shown in Figure 4.9. The Kruskal-Wallis and Mann-Whitney U Tests were also performed but yielded insignificant results. Essentially, users did not spend a significantly different amount of time in top view between mode, day and update cases. There were only slight effects for trial which could be impacted by the amount of QR codes scanned or the length of the trial since this was not normalized for trial length or mean arrival time.

Analysis of side view time yielded no significant effects for original values and natural log transformed values in mixed linear regression models nor Kruskal-Wallis and Mann-Whitney U tests. Therefore users did not spend a significantly different amount of time in side view between mode, day and update cases.

Overall subjects use of top view and side view shows that there were no significant

Table 4.9: Natural Log of Top View Time

Variable	$\hat{\beta}$	SE	Z	p
Intercept	11.47	0.06	198.31	<0.001
Trial 1	-0.08	0.04	-2.16	0.031
Trial 2	0.04	0.04	1.02	0.310
Trial 3	0.08	0.04	2.10	0.035
Trial 4	0.03	0.04	0.74	0.461
Trial 5	-0.08	0.04	-2.27	0.023

differences in which view subjects utilized during a trial. Considering side view use was insignificant, the natural log of top view time model has minuscule differences in percent usage between trials. Therefore, it appears that the use of these two views was balanced throughout the trials, with mode, inclusion of an update, the trial, the day or the order not changing the users' general use habits.

Arrival time was fitted to a linear regression with order and significant effects of mode with an interaction effect of day-order. These can be seen in Table 4.10 with estimates in milliseconds. Users spent 4.026 less seconds than the mean in computer-select and 4.756 more seconds in the even group. Users spent more time in each trial in user-choice mode and in the even order group. The interaction effect estimates that even on day 1 and odd on day 2 spent 2.91 less seconds to finish the trial than the mean, while even on day 2 and odd on day 1 spent 2.91 more seconds to finish the trial than the mean. Whereas the even order group continued to push the boundaries of the time limit, the odd group was much faster on day 2 and this accounted for the interaction effect.

Overall, the arrival time statistics demonstrate that user's spent more time completing trials in user-choice mode, even though it was shown earlier that this did not lead to higher point totals on day 1 and comparatively lower gains on day 2. This increase in arrival time is most likely due to the additional interactions required by user-choice mode to operate. The order effect was significant in this case and there was shown to be a significant interaction effect with order-day. This is also shown in plan time and supports the idea that subjects in the odd group were more decisive in day 2 of the experiment by spending less time planning and completing trials, while

Table 4.10: Arrival Time

Variable	$\hat{\beta}$	SE	Z	p
Intercept	157343	2820	55.79	<0.001
Mode	-4026	1472	-2.74	0.006
Order	4756	2819	1.69	0.092
Order*Day	-2910	1467	-1.98	0.047

the even group stayed mostly consistent. This could be that the odd group was just more efficient in their route planning and execution in accomplishing the same task. This suggests that even subjects continued to exhaust all options to maximize their point totals on day 2 and thus the time did not change. Odd subjects kept similar strategies and this suggests that they became more efficient and spent less time enacting them.

To look at the map view time more closely, top view time was normalized by the arrival time to limit variance due to trial time. This permitted the analysis of the percentage of time users spent in the top view relative to the trial duration. There was a significant covariate effect of this ratio to points collected after penalties with 0.0034 seconds per point after penalties added to the amount of time spent in top view relative to the length of the trial (Figure 4-2). Therefore users spent a greater proportion of the time in each trial in top view when they received more points and thus had to access the QR code scanner button more frequently. It also probably means that users had the route mentally mapped out, as shown in their subjective statements about route change. This could result in subjects using the phone less to navigate and thus the phone was mostly in top view.

QR scan time was also transformed by natural log and yielded significant effects of order, update, day and Trial 2 with an interaction effect between day-mode as shown in Table 4.11. Therefore users spent more time scanning QR codes if they were in the even ordered group of 0.11 percent, on day 2 of 0.07 percent, and there was no update of 0.08 percent. This directly corresponds to the results of total points collected with and without penalties. The improvement in use of user-choice mode is shown in the interaction effect with an estimate of 0.06 percent more time above the

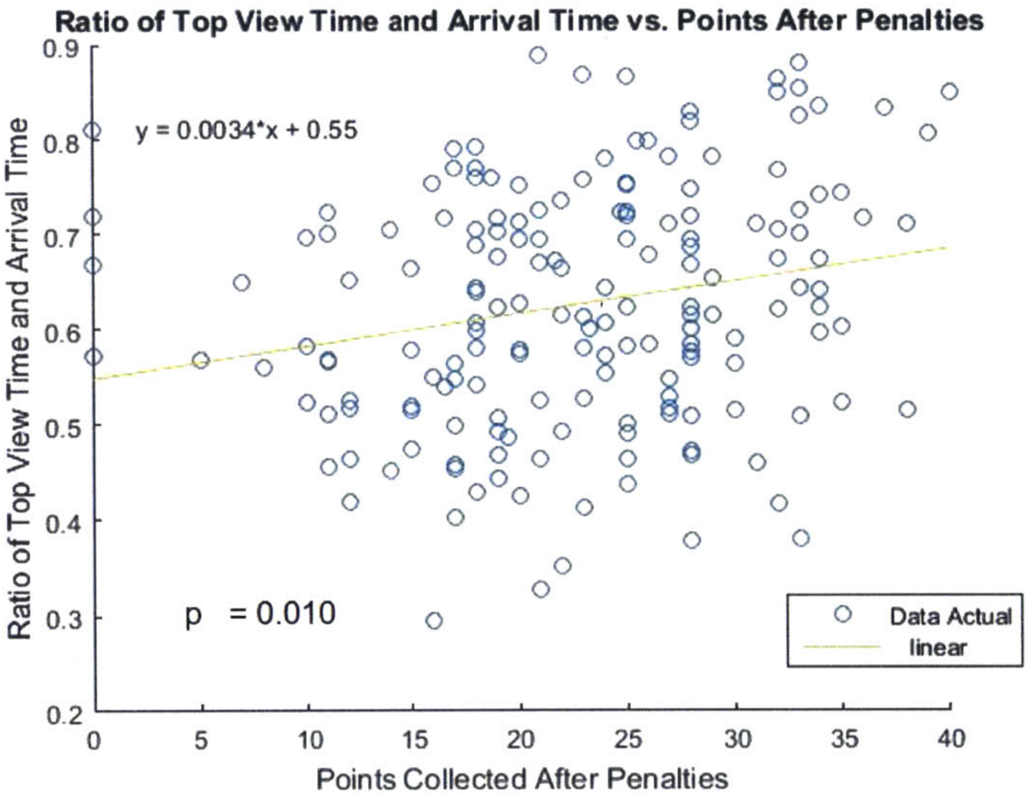


Figure 4-2: Ratio Between Top View and Arrival Time Linear Regression Results

Table 4.11: Natural Log of QR Code Scanner Time

Variable	$\hat{\beta}$	SE	Z	p
Intercept	10.26	0.06	180.72	<0.001
Update	0.08	0.02	4.49	<0.001
Day	-0.07	0.02	-3.92	<0.001
Order	0.11	0.06	1.99	0.047
Trial 1	-0.34	0.04	-0.86	0.391
Trial 2	0.11	0.04	2.81	0.005
Trial 3	0.08	0.04	1.84	0.065
Trial 4	-0.001	0.04	-0.04	0.971
Trial 5	0.02	0.04	0.48	0.629
Mode*Day	0.06	0.02	3.55	<0.001

mean in computer-select on day 1 and user-choice on day 2. In addition, it appears there was a tiny benefit in scan time for Trial 2. In general this variable simply supports previous findings especially since the model is similar to the model for total points collected before penalties except for order. This variable is not particularly operationally relevant beyond the support.

There was no support for either STH1 or STH2 based on this analysis. However, the timing aspect of the performance metrics does support the score conclusions. Further analysis of the impact of plan time by constraining that parameter would provide more information about its impact on mode preference and performance. An additional look at arrival time, by providing a reward for arriving early would also be interesting for user strategy.

4.2.3 Strategy

Route changes was initially analyzed using a linear regression model. The initial model found significant effects for mode and update, but failed the Kolmogorov-Smirnov normality test and Levene's test for constant variance. Looking closer at the residuals, it was determined that they fit into three bands based on mode and update. Using this knowledge, two more models were attempted looking at the two modes separately. There was found to be no difference between the update groups within user-choice but there was a significant effect in computer-select as shown in

Table 4.12: Route Changes Model for Computer-select Only

Variable	$\hat{\beta}$	SE	Z	p
Intercept	5.52	0.36	15.35	<0.001
Update	-0.73	0.23	-3.15	0.002

Table 4.12. This shows that the model predicts user's to have 0.73 more route changes than the mean when an update occurs. A Mann-Whitney U test showed that the two modes were significantly different ($p < 0.001$) as shown in Figure 4-3. Essentially, there were more route changes when there was an update for computer-select only. It also is very apparent that subjects repeatedly chose to not change their route regardless of updates in the user-choice mode, because the mean was below one for both cases. This means that even with an update, subjects did not re-plan in the system because the additional effort was unnecessary to their physical navigation.

Overall, this analysis supports the claim that subjects did use the two modes differently. The additional action of requiring the user to re-select their route to update their path in user-choice was considered tedious regardless of whether the updated path was to correct for the user's deviation, slow pace or possibility of more points. The time sensitivity of the task pushed the subjects to just ignore the app and its updates. This is reflected in the data being skewed toward zero for user-choice. In the computer-select mode, the user had to accept every updated path and thus we see a significant difference in the number of route changes with an update which is fairly reasonable since this would cause a subject to deviate more. In addition, there were shown to be dependencies on order and these appear to signal the adoption of route change policies by the two groups with the odd group adopting this policy quicker, which would result in them having fewer number of route changes.

System interactions was fitted to a mixed linear regression model with significant effects for mode, update and order with an interaction effect between mode-update as shown in Table 4.13. This model shows that users interacted with the system more when there was an update by 2.05 interactions, in user-choice mode by 3.32 interactions and in the even order group by 1.94 interactions. There was also an interaction effect between mode-update with an increase of 1.55 interactions with an

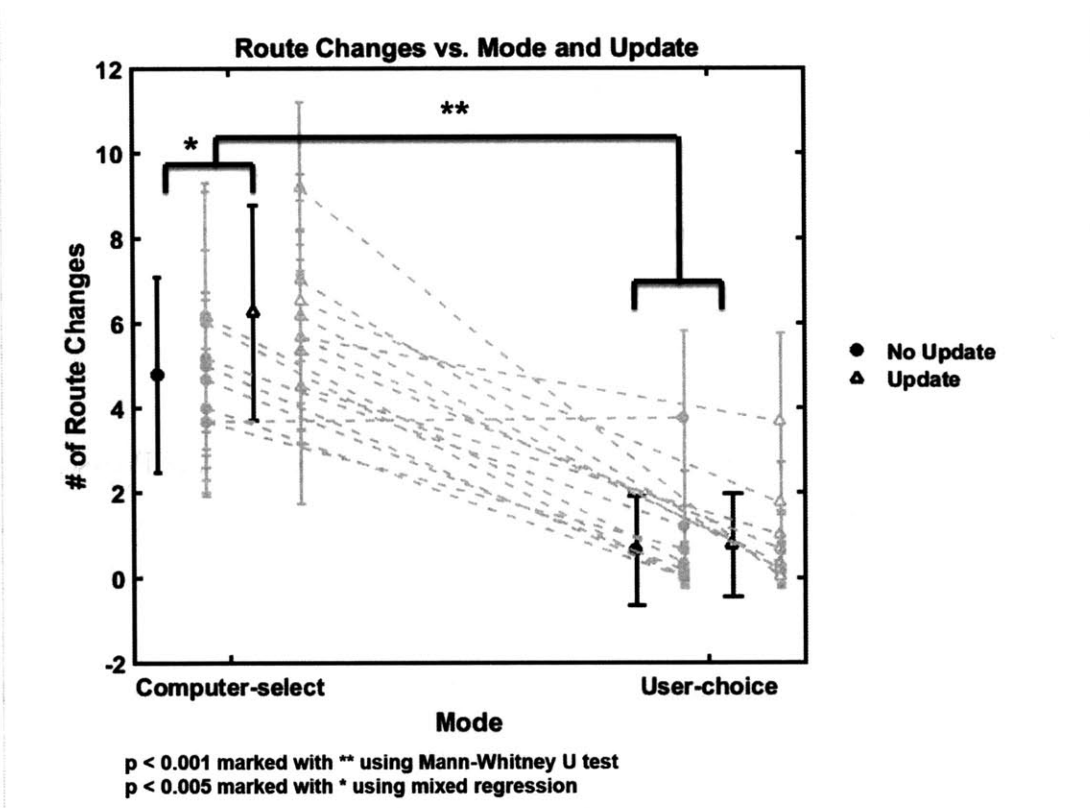


Figure 4-3: Route Changes Graphical Representation with grey lines representing subject means and standard deviations while black lines are the group means and standard deviations

Table 4.13: System Interactions

Variable	$\hat{\beta}$	SE	Z	p
Intercept	22.40	0.54	41.52	<0.001
Order	1.94	0.54	3.60	<0.001
Update	-2.05	0.45	-4.58	<0.001
Mode	-3.32	0.45	-7.39	<0.001
Mode*Update	1.55	0.45	3.46	0.001

update in user-choice and no update in computer-select than the mean.

Overall this analysis indicate that users interacting with the system more frequently in user-choice despite less route planning and changes. There was also more interaction with an update which follows the expectation. Based on these two findings, it makes sense that update in user-choice caused a majority of these differences. Finally there were was more interactions with the users in the even group, which could be related to the longer time spent completing a trial and using the QR code scanner.

Trials which subjects reported strategy changes were analyzed using Pearson Chi-squared tests. There was shown to be dependence with day, order and interaction effects between mode-day, mode-order, and order-day. These results can be seen in Figure 4-4. These results support the idea that subjects better understood the system on day 2 since there were less strategy changes. Most of the changes were made by the even group and split between the two days evenly. This could be a reflection of the 10 trials that froze on these subjects and their continued work to develop better strategies to get more points. Since even subjects saw computer-select second, more changes were made on day 1 in that mode. This could reflect the user's attempts to modify their use of that mode to match the initial route planning of user-choice.

The subjective strategy responses were grouped (Appendix T) by their theme into the following categories: aggression, SA task, update planning, alerts, route planning, willingness to deviate, deviation planning, and memorization of the map.

These comments speak to the mindset and motivation of participants during the study. Generally subjects were looking at ways to better utilize the system to successfully improve their performance. Solutions included working harder, memorizing the map and route ahead of time, ignoring alerts, full route planning prior to the trial, using the given route, going for high point QR codes and deviating as necessary while also keeping their head up more to improve SA.

SRH1 was fully supported by this analysis, while SRH2 was supported for computer-select only. There was also support for SSH1 and SSH2, while no support for SCH1. These results show that subjects were manipulating their strategy independent of floor closure events and were willing to accept route changes as long as they were not required to inform the system beyond their actions.

4.3 Situation Awareness

This section will look primarily at the subject's performance in the secondary task of situation awareness. The focus is on the items from the main SA survey (see Appendix N). Table 4.14 shows the general trends of the subjects overall SA.

Composite SA was fitted to a mixed linear regression model with a significant effect for day and an interaction effect between mode-update as shown in Table 4.15. This model shows that users had better SA on day 2 by 0.97 points above the mean. In addition there was an interaction effect where subjects had worse SA with an update in computer-select and without an update in user-choice of 0.78 points below the mean than the other two conditions.

The improvement in composite SA coincides with the user's increased familiarity with the system, which allowed the subject to have more attention to be directed to the environment and explains the improvement on day 2. The interaction effect could possibly be explained by the user improving SA in user-choice with an update due to increased knowledge gained from the app if they were indeed ignoring the app's updates. This forced the user to look at the app and select a new route which could reorient them. While the update in computer-select would not have had the same effect because the user was only required to acknowledge the update. There was no main effect for mode or update when pooled because these changes averaged out.

The delta between a subject's recalled and actual initial route points were initially

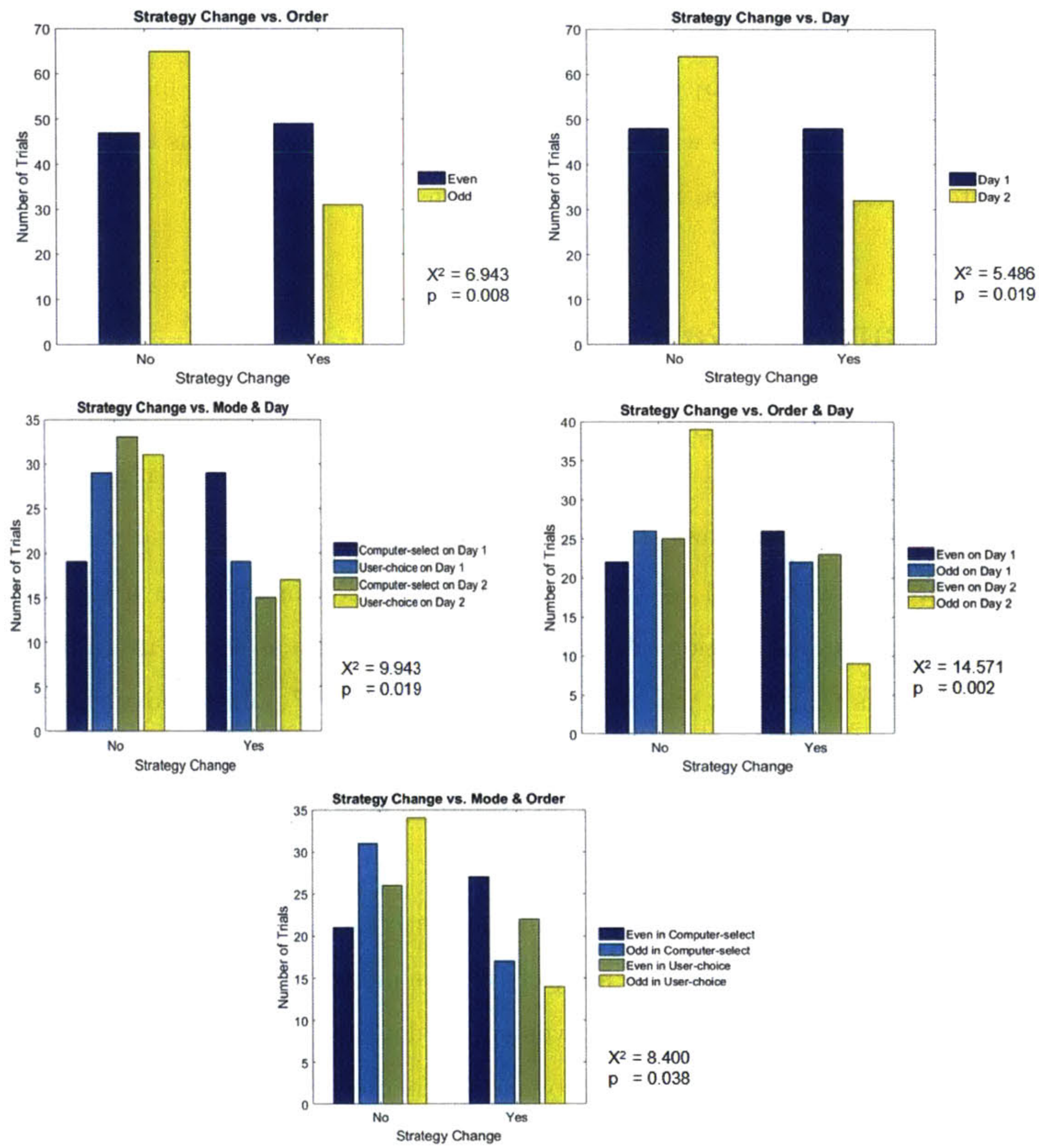


Figure 4-4: Strategy Change Contingency Table Results

Table 4.14: General SA During the Experiment

	Fraction	Percent
Start	150/192	78%
Goal	171/192	89%
Trials that Characters were seen	176/192	92%
Characters correctly seen	277/292	95%
Correctly Placed on Floor	239/292	82%
Correctly Placed at Location	175/292	60%
Floors Correctly Recorded Visiting	802/882	91%
	Total	Total/Trial
Total Characters Seen	292	~1.5
Different Floors with QR Codes Scanned	882	~4.6

Table 4.15: Composite SA

Variable	$\hat{\beta}$	SE	Z	p
Intercept	11.47	0.60	19.00	<0.001
Day	-0.97	0.28	-3.51	<0.001
Mode*Update	0.78	0.28	2.83	0.005

fitted to a linear model, but the model failed Kolmogorov-Smirnov normalcy test and Levene's test of constant variance. Therefore this variable was analyzed using Pearson Chi-squared tests.

The delta values were looked at as absolute values since the magnitude of error was more pertinent. There was shown to be dependence with order and trial, and interaction effects between order-day, update-order, and mode-order. These results can be seen in Figure 4-5. These results show that the odd group was more accurate when accounting for order, day and update. There was also an effect for Trial 2 with subjects being inaccurate on more than half of their trials in this scenario.

In general, subjects did a very good job of being within a point of the initial route they were given. Trial 2 apparently disoriented users more but it is hard to draw any additional meaning from this value. This could also be impacted by the length of time spent planning and analyzing the three options in user-choice, but there was only the order-mode interaction effect.

Subjects' correctness for start locations was analyzed using Pearson Chi-squared tests. There was shown to be dependence with day and trial, and interaction effects between mode-day, and update-day. These results can be seen in Figure 4-6. These figures shows that subjects improved on day 2 and most successfully in user-choice and no update.

As shown, it appears that trials 3, 4 and 5 were more difficult to recall the starting location. This could be due to these trials having start and end locations in the middle of the floors, which may be to harder to recall than knowing the extremes. Subjects appeared to improve on day 2 in both modes and with or without an update.

Subjects correctness' for goal locations was analyzed using Pearson Chi-squared tests. There was shown to be dependence with mode, and interaction effects between update-day. These results can be seen in Figure 4-7. This figure shows subjects improved on day 2 with an update and performed better in user choice. This suggests that the update disoriented the users on day 1, but their improvement in system use allowed them to fix the issue.

The goal location was planned as a check for SA with the expectation that there

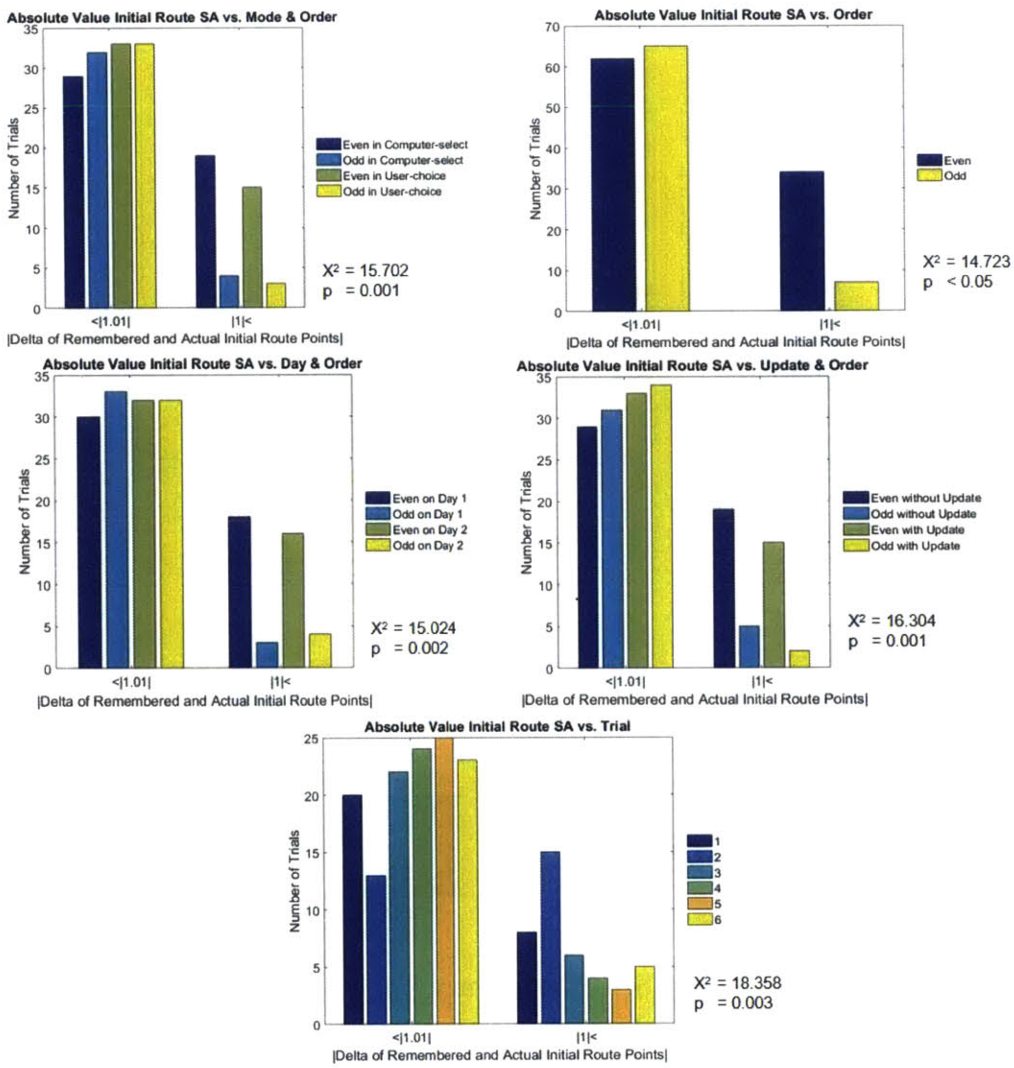


Figure 4-5: Absolute Value of Delta Recalled and Actual Initial Route Points Contingency Table Results without One Subject

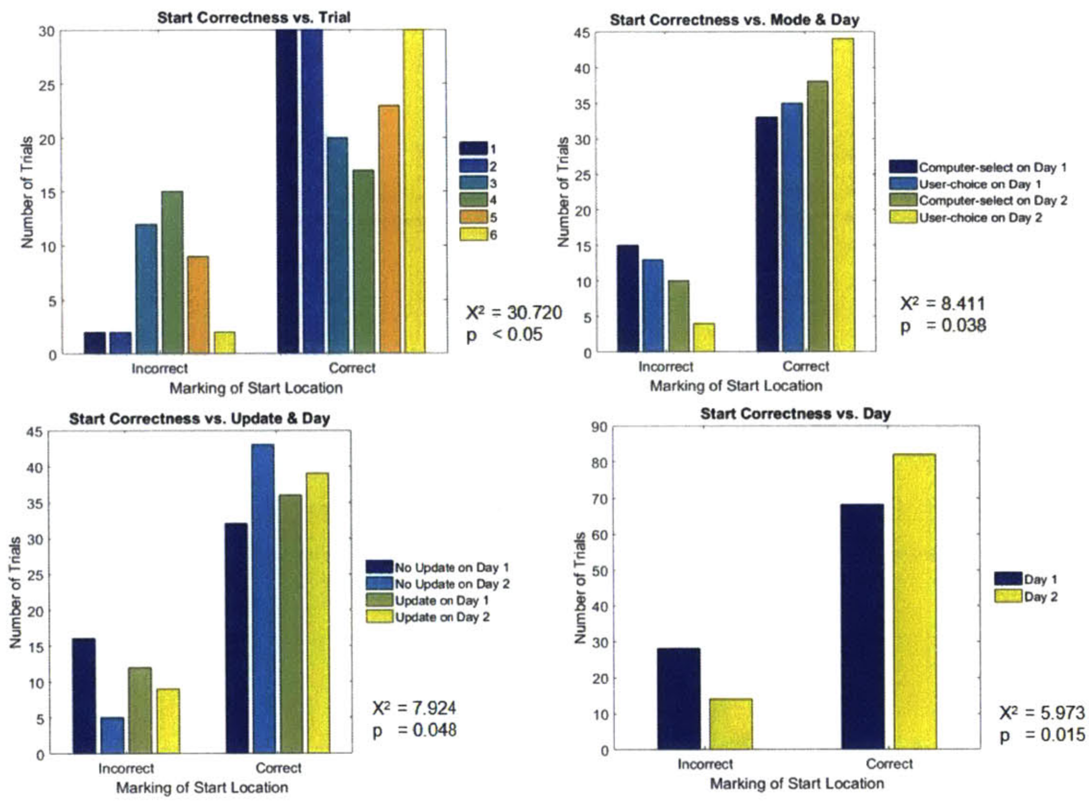


Figure 4-6: Start SA Contingency Table Results

would be no incorrect responses since subjects completed the survey while standing at the goal location. The fact that 11 percent of the responses were incorrect was surprising. This implies the subjects had no clue where they actually were. Most of these cases were on day 1 and with an update which could have caused the additional stress. In addition most of the incorrect responses were in computer-select which required less knowledge of the environment and experienced less planning time.

Subjects' correctness for marking the floor location of characters was analyzed using Pearson Chi-squared tests. There was shown to be dependence with update. This result can be seen in Figure 4-8. This figure shows mixed results with update. This could be confounded by the number of floors visited by a subject. Therefore, the key takeaway is that remembering the characters, let alone the correct floor was a difficult task. Even though subjects scanned QR codes on about 4.5 floors per trial, almost two-thirds of their responses were one or less.

Subjects' correctness for marking floors that they scanned QR codes was normalized by total number of floors visited and analyzed using a mixed linear regression model but the model failed the normality assumption. Using the Mann-Whitney U test, the ratio of correctly visited floors was found be significant for day ($p= 0.035$). The Kruskal-Wallis test for more than two treatments showed trial to be significant ($p = 0.018$), with only the comparison between trial 1 and 5 being significant ($p = 0.028$) for the Dwass-Stell-Chritchlow-Fligner Test for all pairwise comparisons (Figure 4-9). This suggests that subjects did better at this task on day 2 and at trial 1 as compared to trial 5 which had a large variation. This tells that the use of the extremities of the map as start and end locations for trial 1 made it relatively easy for the subjects to track their own movements, while trial 5 was the most disorienting. There were a total of 18 occurrences of user's scanning QR codes worth zero points.

SAH1 and SAH2 are only partially supported in that mode and update impact on SA were connected through an interaction. The larger impact was the familiarity of the task over time.

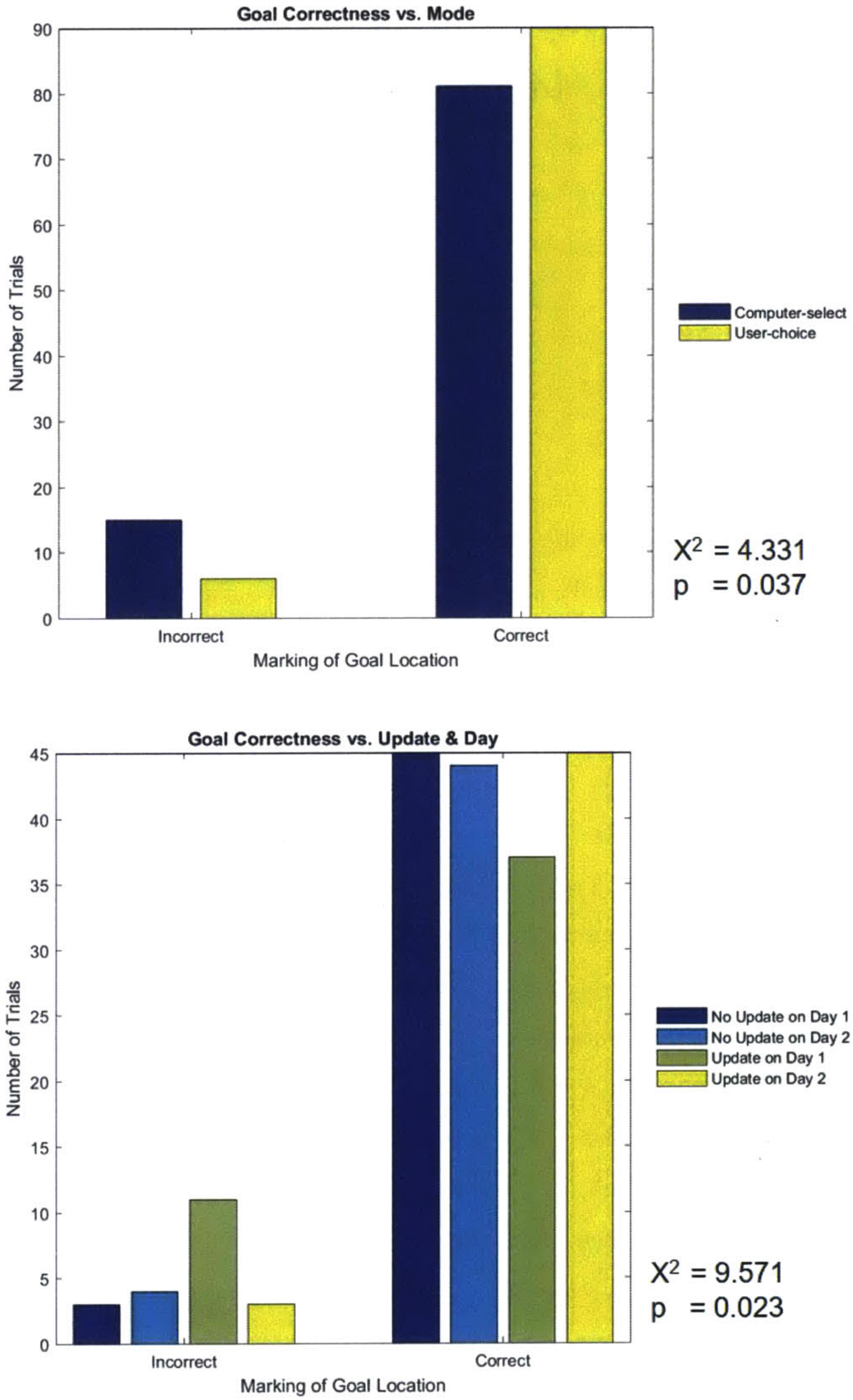


Figure 4-7: Goal SA Contingency Table Results

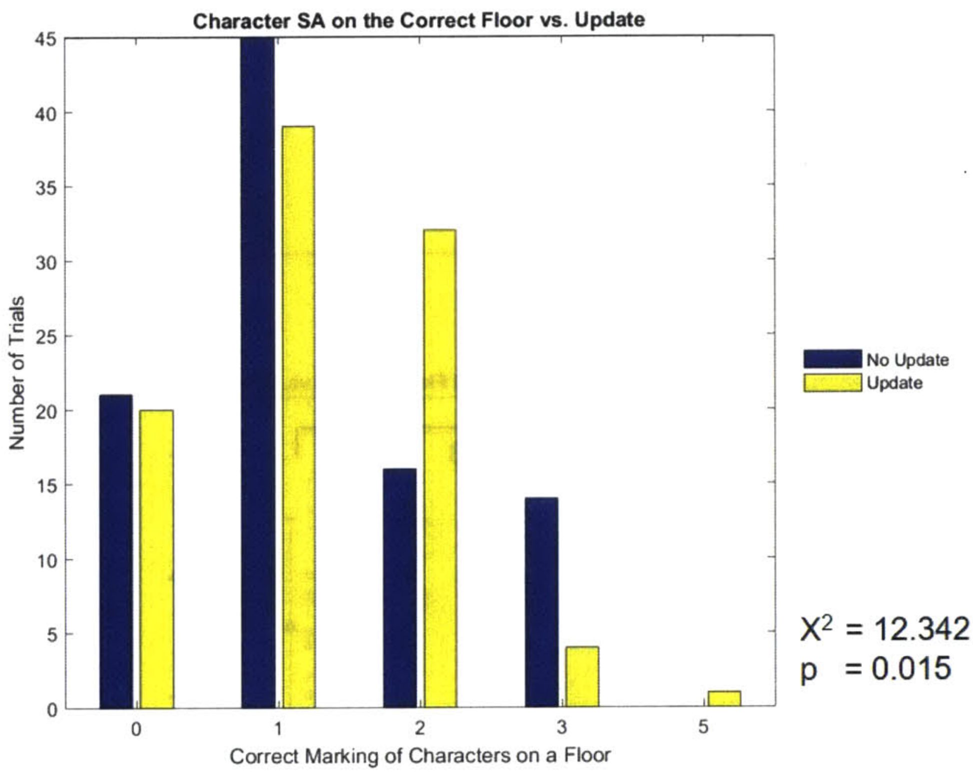


Figure 4-8: Character SA on Correct Floor Contingency Table Results

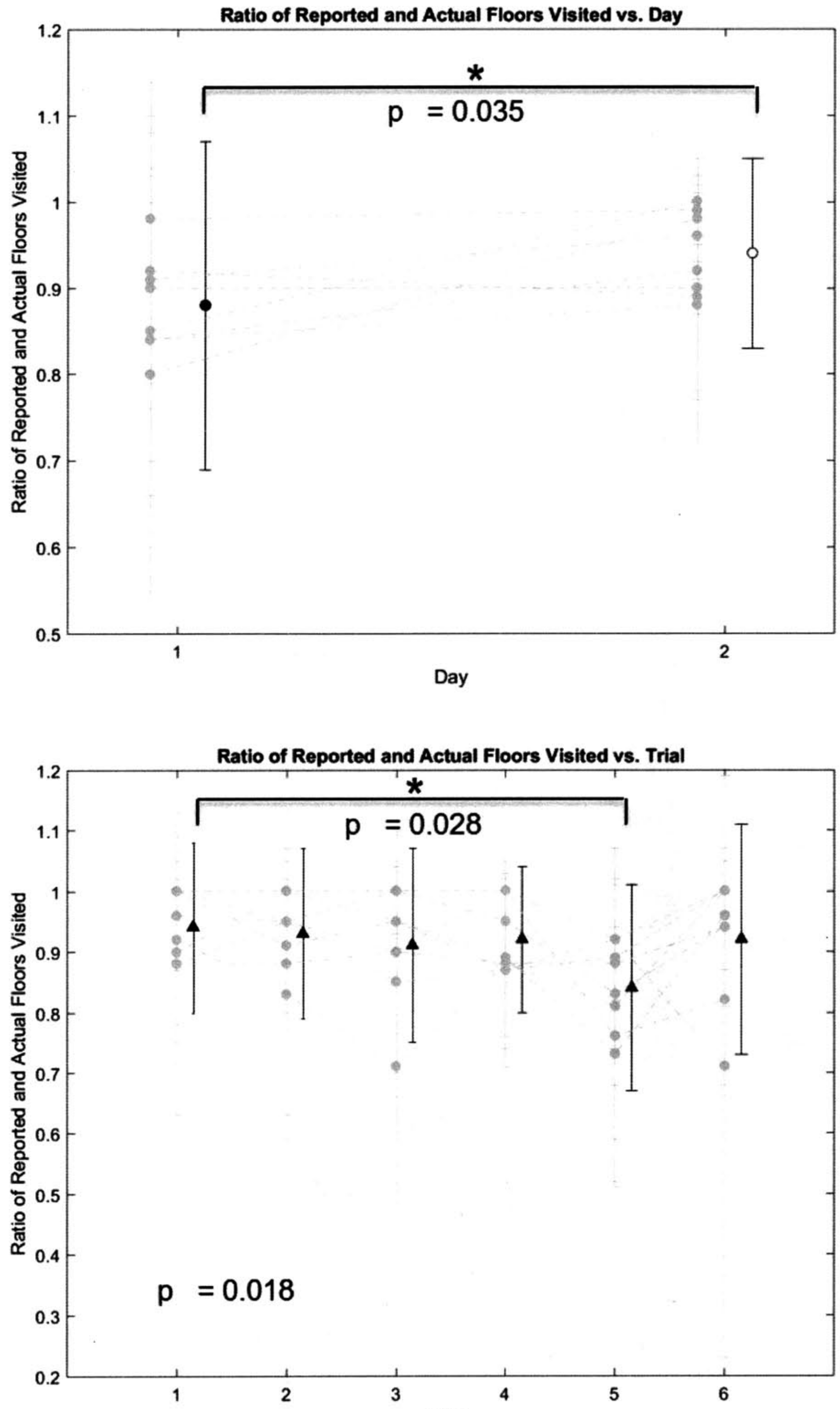


Figure 4-9: Ratio of Reported and Actual Floors Visited with grey lines representing subject means and standard deviations while black lines are the group means and standard deviations

Table 4.16: General Workload During the Experiment

		Mean	SD
Mental	Raw	62.422	14.392
	Weighted	16.245	5.22
Physical	Raw	49.219	23.046
	Weighted	7.271	9.097
Temporal	Raw	66.172	23.498
	Weighted	17.166	10.26
Performance	Raw	44.609	22.805
	Weighted	6.933	5.168
Effort	Raw	61.562	17.477
	Weighted	8.984	6.5
Frustration	Raw	43.438	24.117
	Weighted	5.703	8.538

Values range from 1 (Strongly Disagree) to 5 (Strongly Agree)

4.4 Workload

This section will look primarily at the subject's workload during each trial. Table 4.16 shows the general workload during the experiment from the NASA TLX test in its raw and weighted components.

Composite TLX workload was originally fitted to a mixed linear regression model, but this model failed Levene's Test for constant variance. Therefore, TLX composite was analyzed using Pearson Chi-squared tests. There was shown to be dependence with day and update. These results can be seen in Figure 4-10. This figure demonstrates that users had higher workload levels with an update and on day 1 which corresponds to the subject learning curve displayed by the other variables and the disorienting effect of an update.

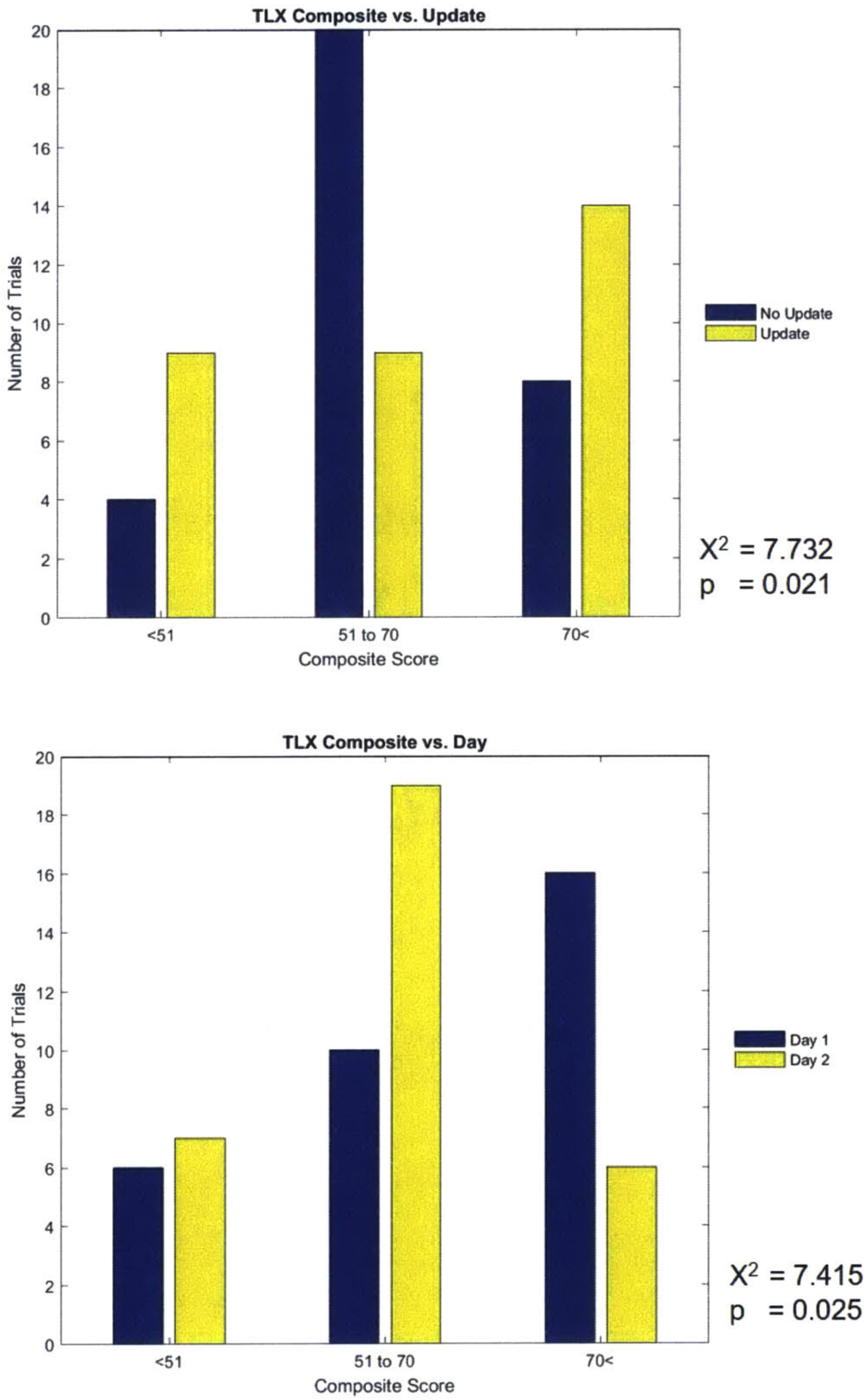


Figure 4-10: NASA TLX Composite Contingency Table Results

Subjects were given additional subjective questions on workload that were filled out every trial to include: physical and mental demands of the task, and how much the app reduces these demands in their current mode.

The physical demand of subjects was analyzed using Pearson Chi-squared tests. There was shown to be dependence with day and order, and interaction effects between mode-day, order-update, order-mode, and order-day. These results can be seen in Figure 4-11.

The key takeaway from this analysis is that a majority of the subject's responses were that the physical demand of the task was moderate with very few subjects reporting high demands in this task. This supports the utilization of the military population to minimize physical effects.

The reduction of physical exertion due to the app was analyzed using Pearson Chi-squared tests. There was shown to be dependence with day and order. These results can be seen in Figure 4-12.

The basic requirements of the protocol required several stair changes and the app was not primarily designed to mitigate the physical demand. Therefore the low values of physical reduction due to the app align, though subjects felt more satisfied on day 2.

The reduction of mental exertion due to the app was analyzed using Pearson Chi-squared tests. There was shown to be dependence with mode. This result can be seen in Figure 4-13. This shows that subjects felt that user-choice contributed less to the reduction of the mental part of the task.

The dependence of mode contains a majority of responses as moderate. The mode groups only deviate on the low responses with many more user-choice trials qualifying. This supports that subjects required more experience to operate user-choice effectively and thus more subjects felt the system was not lowering their mental demand.

The subjects fatigue was analyzed using Pearson Chi-squared tests. There was shown to be dependence with order. This result can be seen in Figure 4-14.

Most of the responses fall in the middle band. Though more even group subjects responded with more fatigue this could be due to the longer trials and more ambitious

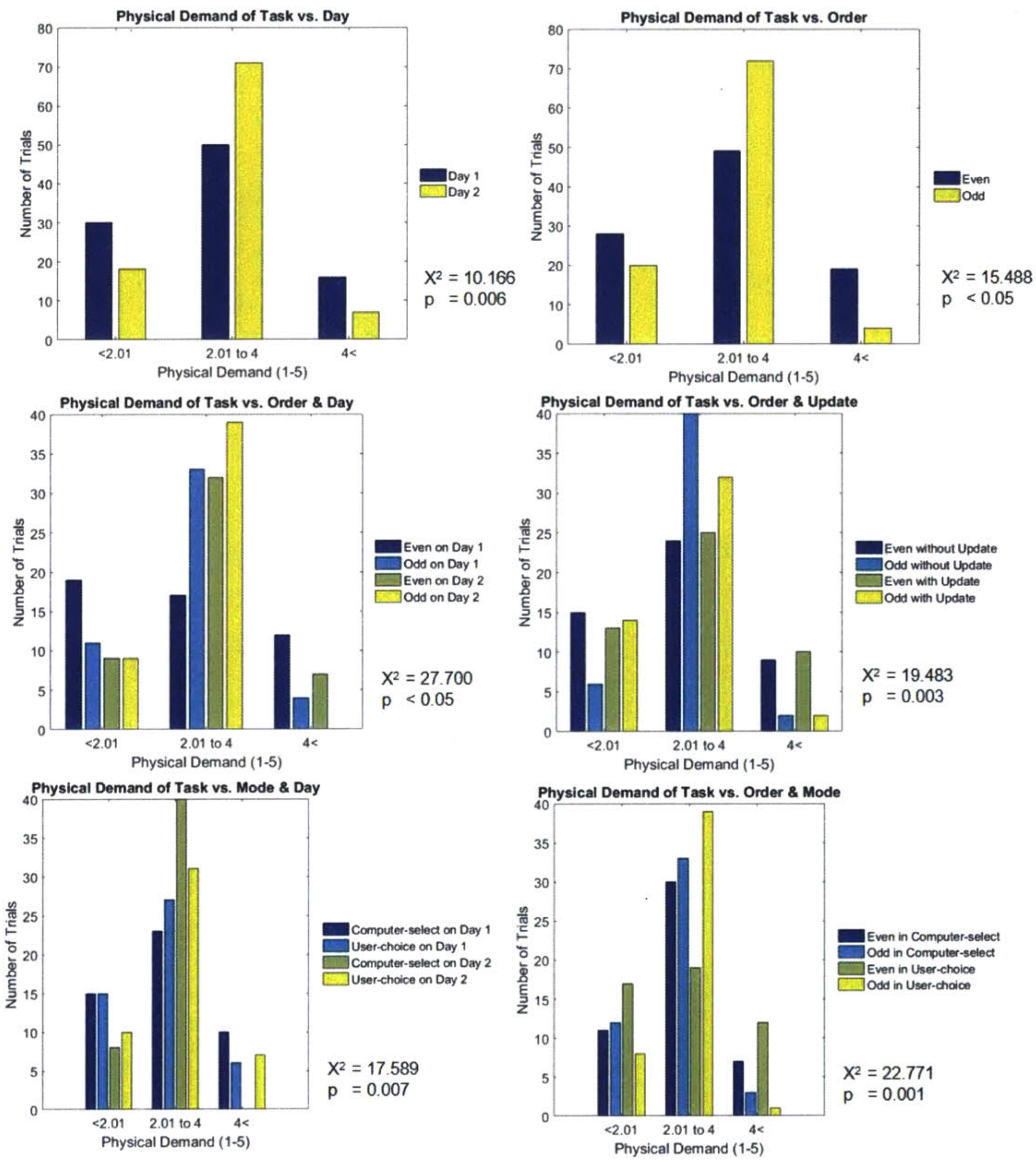


Figure 4-11: Physical Demand Contingency Table Results

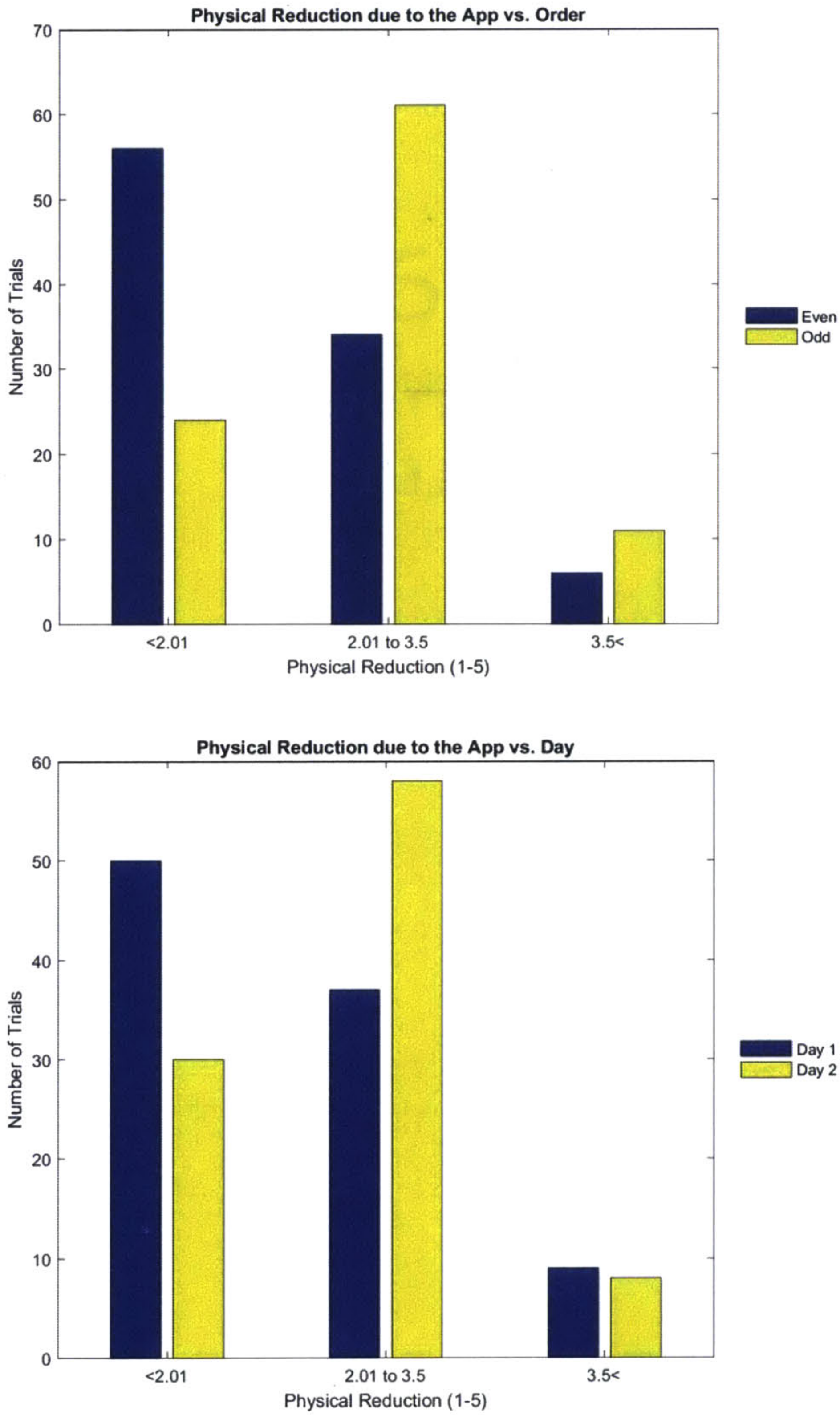


Figure 4-12: Physical Reduction Contingency Table Results

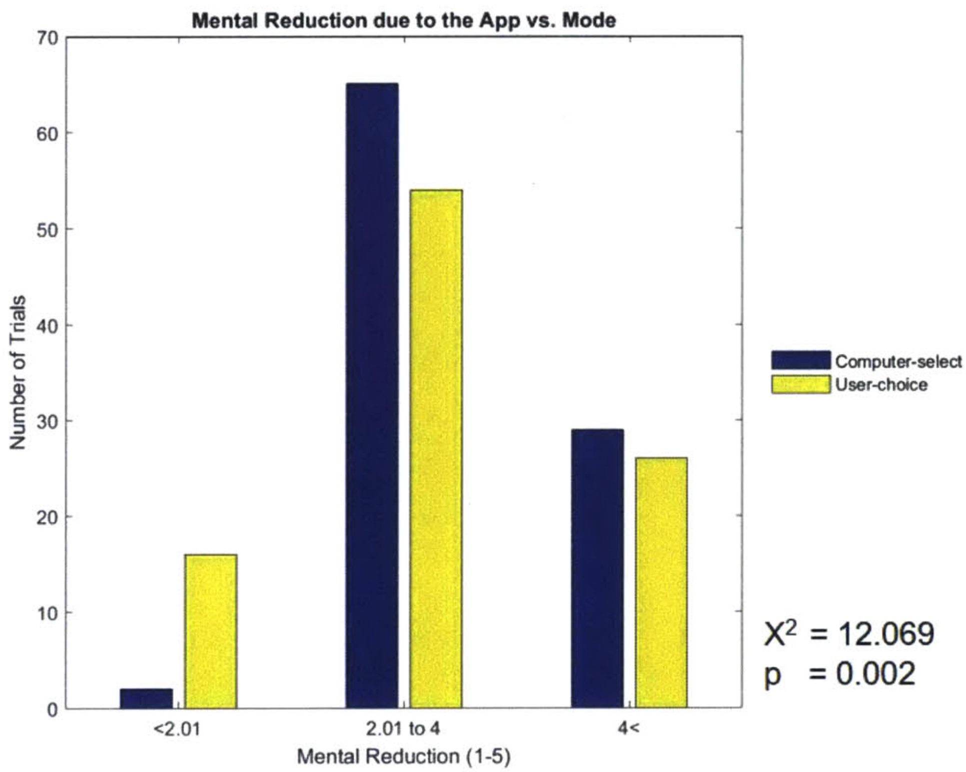


Figure 4-13: Mental Reduction Contingency Table Results

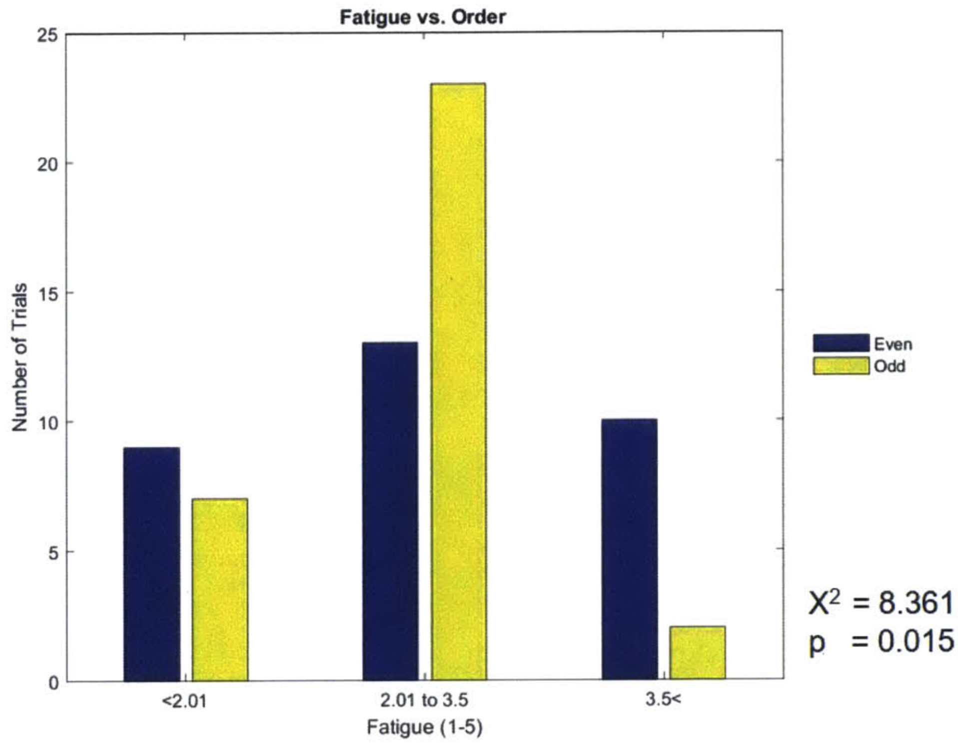


Figure 4-14: Fatigue Contingency Table Results

routes they completed.

There was no support for WH1 and WH2 as the day of the experiment was more important to determine the user's workload. WH3 was supported by the experiment for the TLX composite. WH4 was not supported by the analysis as it was partially controlled by the subject order and within subject design.

4.5 Trust

This section will look primarily at the subject's trust in the app between the eight parameters. Table 4.17 summarizes the the group means and standard deviation for the eight trust questions.

The trust composite was fitted to a mixed linear regression model with significant effects for mode, day and order as shown in Table 4.18. This model shows that users had more trust on day 2 by 1.25 points above the mean, in computer-select mode

Table 4.17: General Trust of the App

	Mean	SD
System Trust	2.72	0.945
Route Changes Appropriate	2.87	0.751
App Clarity	3.231	0.852
Changes Expected	3.094	0.908
App Refresh	2.934	1.018
Clarity of Route Change Reason	3.443	0.919
Ease of Identifying Changes	3.595	0.663
App Usefulness	3.86	0.882

Values range from 1 (Strongly Disagree) to 5 (Strongly Agree)

Table 4.18: Composite Trust

Variable	$\hat{\beta}$	SE	Z	p
Intercept	25.75	0.72	35.88	<0.001
Day	-1.25	0.38	-3.32	0.001
Mode	0.81	0.38	2.14	0.033
Order	-2.05	0.72	-2.86	0.004

Table 4.19: System Trust

Variable	$\hat{\beta}$	SE	Z	p
Intercept	2.72	0.13	20.58	<0.001
Order	-0.53	0.13	-4.03	<0.001

by 0.81 points above the mean and in the odd order group by 2.05 points above the mean.

The improvement in trust over time supports the work by Lee & See (2004) that as the subject's experience increased, the subject used the system more appropriately [28, 33]. The composite also shows that subjects trusted the computer-select mode more, which could stem from them using that mode more appropriately earlier based on the consistency in their use of the system. This also demonstrates that subjects in the odd group generally had more trust. The composite trust of the even group was negatively impacted by trials where the app crashed since 10 of the 12 were in the even group. This would significantly affect the subjects trust. This is even reported in one qualitative response for mode preference. Each component of the trust composite was analyzed individually to see if it was significant by itself.

Overall system trust was fitted to a linear regression model with a significant effect for order as shown in Table 4.19. This model shows that the odd group had more system trust by 0.53 points above the mean. This shows that overall system trust was a key component in the trust composite model and this effect could be impacted by the trial failures.

Trusting the app's route changes was fitted to a linear regression model with a significant effect for day as shown in Table 4.20. This model shows that users trusted the route changes more on day 2 by 0.19 points above the mean. This shows that

Table 4.20: Trust of Route Changes

Variable	$\hat{\beta}$	SE	Z	p
Intercept	2.87	0.13	22.76	<0.001
Day	-0.19	0.08	-2.31	0.021

Table 4.21: App Refresh Reliability

Variable	$\hat{\beta}$	SE	Z	p
Intercept	2.93	0.21	13.98	<0.001
Day	-0.22	0.07	-3.00	0.003
Mode	0.27	0.07	3.65	<0.001
Order	-0.50	0.21	-2.39	0.017

the subject's ability to appropriately trust the system was partially dependent on understanding the system's route changes.

Trusting the app being up to date was fitted to a linear regression model with significant effects for mode, day and order as shown in Table 4.21. This model shows that subjects had better trust of the app's refresh capabilities on day 2 by 0.22 points above the mean, and in computer-select by 0.27 points above the mean. In addition the odd group had more trust by 0.5 points above the mean. This model aligns with the trust composite model by supporting the same claims. This demonstrates that the ability of the app to track the user was vital to their overall trust. Since the trend is similar, this suggests that subjects development of an expectation for the app refreshing became more consistent with the device through use.

Understanding of the app's reason for route changes was analyzed using Pearson Chi-squared tests. There was shown to be dependence with day. This result can be seen in Figure 4-15. This supports the relationship with the trusting of route changes with subjects reporting a better understanding of the reasons for these changes on day 2.

The ease of identifying changes in the navigation was fitted to a linear regression model with a significant effect for order as shown in Table 4.22. This model shows that the odd group thought it was easier to discern changes in the route on the map by 0.28 points above the mean. This model shows that one reason the difference

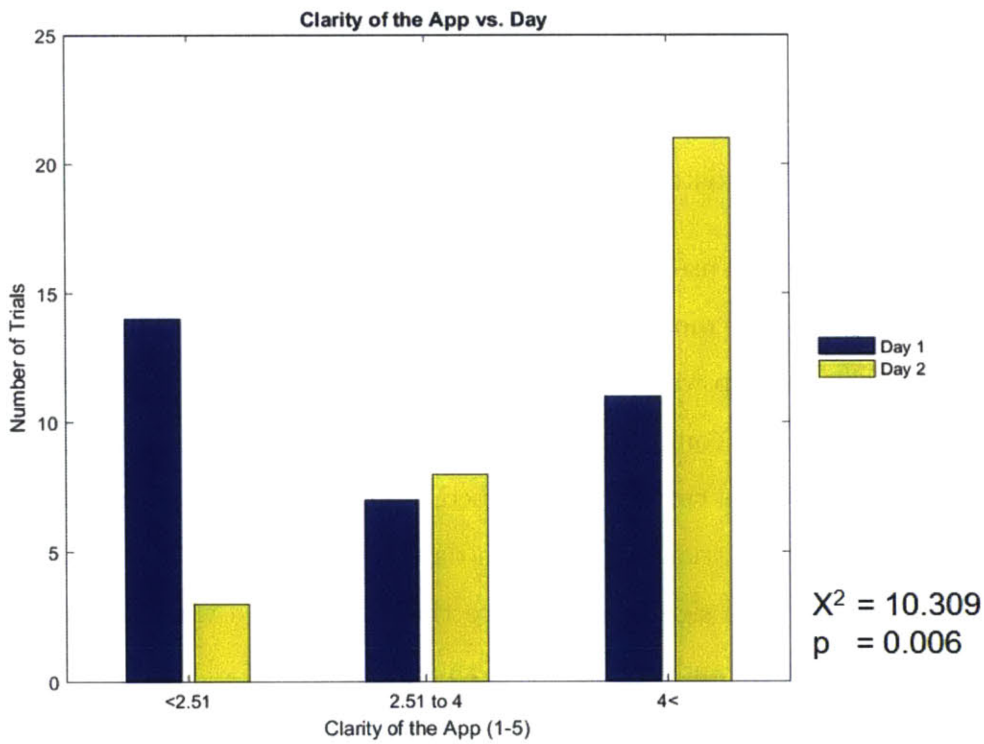


Figure 4-15: Clarity of App Actions Contingency Table Results

Table 4.22: Ease of Identifying Route Changes

Variable	$\hat{\beta}$	SE	Z	p
Intercept	3.60	0.10	37.41	<0.001
Order	-0.28	0.10	-2.94	0.003

Table 4.23: Usefulness of the App

Variable	$\hat{\beta}$	SE	Z	p
Intercept	3.86	0.22	17.73	<0.001
Mode	0.19	0.07	2.52	0.012
Day	-0.21	0.07	-2.76	0.006

between the two groups existed was the ability to identify changes in the app.

The subjects perceived usefulness of the app was fitted to a linear regression model with significant effects for mode and day as shown in Table 4.23. This model shows that users thought the app was more useful in computer-select of 0.19 and on day of 0.21 above the mean. As compared to the other trust metrics, the usefulness of the app was rated highest. As the subjects gained a more appropriate understanding of how to use the app, their trust improved on day 2. In addition, the subjects trusted the computer-select mode significantly more than user-choice which corresponds to the user's open responses and utilization of the system.

A correlation analysis was conducted to compare the GPS post-training trust questions to the mean and median of the related app trust questions by subject. There was found to be a significant relationship with the expectation of route changes for both systems as displayed in Figure 4-16. This demonstrates that in only one of eight cases, the subjects trust of GPS was mimicked by this experiment. This seems to support the notion that preconceived opinions of GPS did not impact the experiment.

Overall there was support for both hypotheses (TH1 and TH2) with significantly more trust in computer-select mode and in the overall system on the second experimental day.

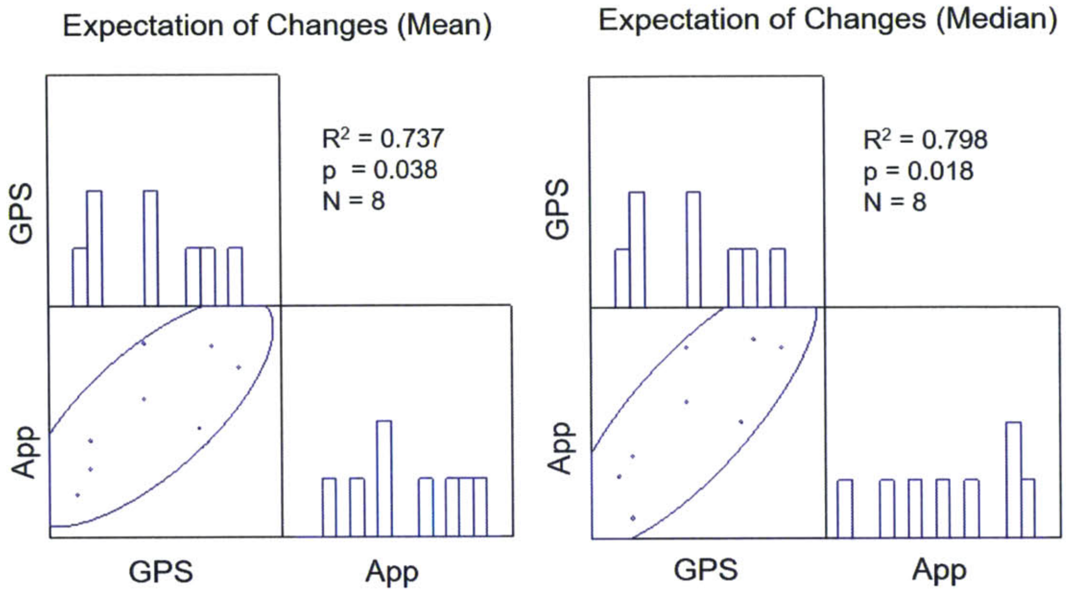


Figure 4-16: Expectation of Changes between GPS and App Correlation Results

Table 4.24: General Scenario Views

	Mean	SD
Task Complexity	3.195	0.791
Task Realism	3.81	0.826
Performance Self-Evaluation	3.575	0.776

Values range from 1 (Strongly Disagree) to 5 (Strongly Agree)

4.6 System Preference

This section will look at subjects evaluations of the task and their performance. Table 4.24 summarizes the means and medians for subjects responses to task complexity and realism, and their self-evaluation.

Subjects performance self-evaluation was analyzed using Pearson Chi-squared tests. There was shown to be dependence with day and order. These results can be seen in Figure 4-17. This shows that generally the even group felt they performed better on the primary and secondary tasks even though there was no support for this claim. In addition subjects felt they performed better on day 2 which corresponds to

Table 4.25: Task Realism

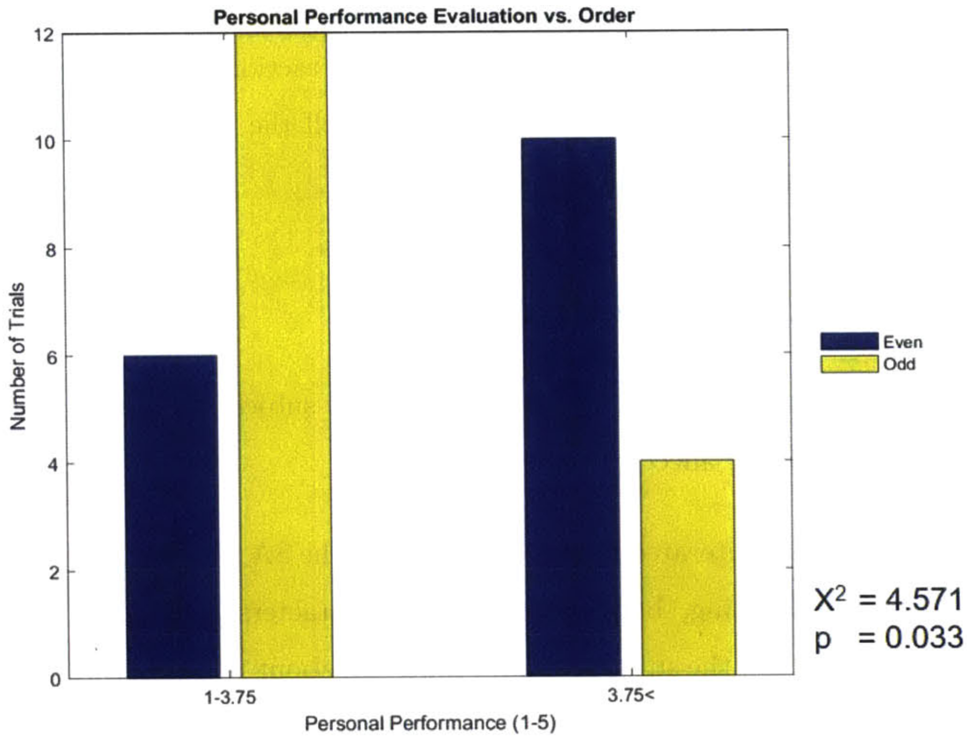
Variable	$\hat{\beta}$	SE	Z	p
Intercept	3.81	0.25	15.11	<0.001
Mode*Day	-0.16	0.07	-2.12	0.034

improved trust, use of user-choice, SA, arriving on time and collecting points.

Each subjects perceived realism of the task was fitted to a mixed linear regression model with a significant interaction effect mode-day as shown in Table 4.25. This model shows that user-choice was reflected as higher realism than computer-select on day 1 by 0.32 points. This effect was reversed on day 2. The intercept of 3.81 out of 5 shows that subjects generally agreed that the scenario was realistic. The interaction effect could be a result of learning from day one to two and bias toward their preferred system. It also aligns with the interaction effect in total points after penalties, thus subjects could be ranking realism based on their performance. The task realism also had an large intercept of 3.81, which supports the claim of this task being realistic.

Subjects preferred the same mode on both days, with six preferring computer-select and two preferring user-choice. It could be relevant that the two subjects that did not experience a trial failure were the two that preferred user-choice. Six of the subjects did select the mode they completed experimental trials for last on the first day. The full responses for their preference can be found in Appendix V. The best and worst features of the app are found in Appendix U.The best parts of the app were the map and in particular the side view, the initial routes, the re-planning, the optional routes, the ability to say no to route changes, the assistance with arriving on time, the mental workload reduction and the WIFI tracking. The major themes of subjects criticism of the app were the WIFI tracking capabilities, lack of user flexibility in path planning, too much user involvement in general, alerts due to tracking, alert frequency, rerouting against the wishes of the user and the side view.

The hypothesis (MH1) that subjects would prefer user-choice was incorrect. However through the explanations for mode selection and the quantitative results it has become clear that the subjects used both modes as a type of user-choice mode. These results demonstrate that users preferred the more difficult and detailed routes of user-



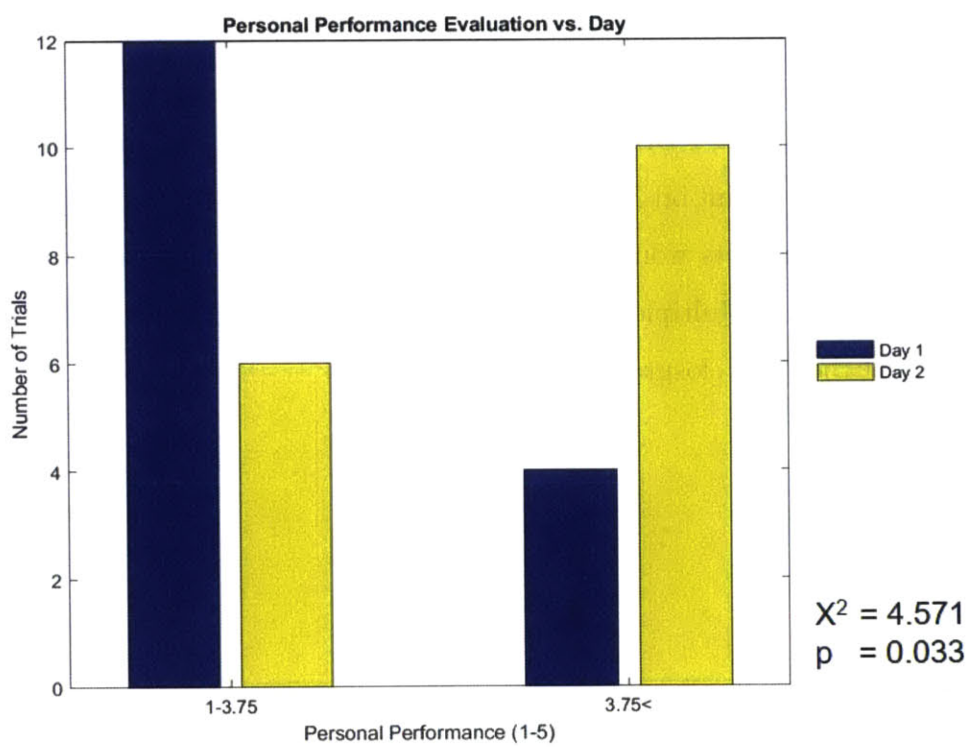


Figure 4-17: Personal Performance Evaluation Contingency Table Results

choice but wanted the system to follow their actions instead of the subject needing to stop and tell the app the planned route. This is reflected in the lack of route changes in user-choice and the amount of system interactions still being higher in user-choice. Therefore, user's were still required to tell the system to not re-route them a comparable amount of times.

4.7 Qualitative Observations

During the experiment, the experimenter followed every subject throughout his entire trial. A couple interesting anecdotes are as follows:

1. Multiple subjects reiterated how the difficulty in the SA task was not in locating them while navigating, but recalling which characters and where they were located. In general, the subjects would complain about remembering one or the other. The speed with which they were trying to gather points was cited as the reason for not clearly remembering. It also did not help that most floors looked the same.

2. In particular in Trial 1 Bravo, subjects would get toward the finish and intend on gathering the point on 37-5, but the floor would close. Instead of proceeding to the finish, subjects would generally go down two flights of stairs to get six additional points. Multiple subjects expressed that this plan was only followed because of the floor closure.

Chapter 5

Conclusion

This chapter will present the contributions and limitations of this thesis as well as future work.

5.1 Thesis Summary

This thesis conducted an indoor search and navigation experiment to determine the appropriate type of automation (adaptive or adaptable) in an uncertain, time-critical scenario applicable to military users and first responders. The specific aims of hypotheses of this thesis were:

1. H1: There is a difference in user performance between adaptive and adaptable automation

   Users performed better in adaptive automation in the primary task overall. This was supported by the findings for PH3 and PH4, which evaluated overall scores with the main effect of mode. These findings showed that computer-select would collect significantly more points than user-choice in the update case and the without update case. These main effect results must be considered in the context of the full model though, because of the significant interaction effect between mode-day with computer-select performing better than user-choice on day 1 and vice versa. The delta first route points results help explain that this interaction was most likely due to the better initial paths being provided in user-choice. This is collaborated by QR code scan time having a similar interaction effect. The findings that looked at main effects for mode and update, of SRH1 and SRH2 on route changes, SSH1 and SSH2 on system interactions and SCH1 on strategy changes, support that the users attempted to apply similar strategies in both modes. This strategy involved deviating from the path to go for more QR codes and gain higher point totals, but only changing the app route, if it was accomplished automatically. The findings of TH1, which evaluated trust and the main effect of mode, support that users trusted the mode that generally performed better and involved the less user planning, which in this case was computer-select. The findings of MH1, which evaluated subjective user preference of mode, support the idea that subjects preferred computer-select which agrees with their performance.

2. H2: There is a difference in performance with and without an update to the scenario

   Users performed better without an update in the primary task overall as supported by the findings for PH1 and PH2, which evaluated the overall scores with the main effect of update. This finding showed that subjects collected more points in both modes in the trials without an update. Subjects' strategy changes were independent of the occurrence of floor closures. The findings for WH3, which evaluated the main effect of update with workload, support the increase in workload for trials with an update.

3. H3: There is an interaction in performance between automation type and update case

   There was not enough evidence to support this claim for the primary task. The findings for SAH1 and SAH2, which evaluated the main effects with SA, support H3. Composite SA improved in computer-select without an update and user-choice with an update. This shows the impact of the mode-update interaction.

Overall this thesis demonstrates that subjects in time-critical, mobile tasks perform better with reduced amounts of control during the task. The study also supports the importance of training on reducing workload, improving SA and increasing user-system trust.

In this experiment, subjects performed better in the primary task in computer-select mode and without an update. Users in user-choice mode tended to have too many alerts for route changes requiring their response and this additional stress resulted in poorer performance even though more planning occurred initially. The additional requirements caused subjects to receive more penalties for arriving late and the accrued penalties were the difference between the two modes. The subjects performed better when the system adjusted its planning by itself in response to the human's actions with computer-select having lower workload and higher trust. The one benefit of the user-choice mode was that subjects developed better SA because their solution to improving their performance was to increase memorization up front and the inclusion of an update actually required them to reorient themselves and thus double check their assumed SA. The SA task coupled with inclusion of updates demonstrated how difficult the task was due to the temporal demands as shown by the SA, workload and trust results. This is best exemplified by the 11 percent failure rate for correctly labeling the goal location while the subjects were standing at the goal. Overall, both modes were used as a type of user-choice mode. Users preferred the more ambitious routes of user-choice but wanted the system to follow their actions instead of slowing down their ability to act.

In addition to the main effect results of mode and update, there was an interesting trend between the two days. In general, subjects improved on the primary and secondary tasks, increased trust and reduced workload between the two days with a majority of these effects resulting from improved performance in the user-choice mode. This demonstrates that the four training trials with one in each condition were not sufficient to fully master the system and understand its reliability. This is also reflected in the subjects self-reporting better performance on the second day.

5.2 Contributions

This thesis contributed the following:

1. Developed an app to perform a navigation experiment
2. Performed a study explaining the applicability of automation to mobile, time-sensitive tasks
3. Developed a list of design recommendations based on this study

Those recommendations for future designs of systems in time-critical, mobile tasks include:

1. Design system to respond to user's actions.
2. Build in flexibility for the users to customize the device. This increases the user's knowledge of the system and improves their expectation for the system's actions. Flexibility should be provided in the planning phase of the task, but not during the time-critical elements. Ideally, flexibility should be at a high enough level that the user can easily understand the impact on the system of selecting a function. For example, providing the user with preset route planning algorithms as in user-choice is very useful, but that initially selected algorithm should not be changed once the trial begins.
3. Increase the amount of training with systems to ensure the users understand the reliability of the system to more appropriately define trust and use strategies. This will also reduce workload and improve secondary tasks.
4. Reduce alert frequency which is a known irritant to user's, exemplified by Clippy [50]. Based on this experiment I would caution against reroute frequencies under five seconds and reroutes that do not change the general landmarks visited or in the case of this experiment QR codes.

5.3 Limitations

The major limitation for this study was the development of the app, because there were 12 trials thrown out due to app freezing. This impacted the affective trust of users and the perceived reliability of the system as evidence by the even group having significantly less trust. The coding of the app was thought to be sufficient at the beginning of the study following numerous user studies and countless trials. Any follow-on studies should update the system software. Additional limitations of the study were the subject population which was limited to military background and male. Other populations and females should be tested in order to generalize the findings of this study. The task itself also provides limitations in that it was constrained to a well-known task environment with a multitude of WIFI APs and indoors. First-responders and military personnel are rarely entering a well known, previously scouted environment and would thus be more reliant on the assistive device to fill in gaps that it could. Constraining plan time for the user would help make the scenario more dynamic and thus realistic.

5.4 Future Work

The research into the use of smart phones in high tempo situations is still sparse and will continue to grow as the military continues to leverage this technology. Expansions on this thesis should look at some of the following:

1. The impact of constraining the amount of planning time on performance would increase the temporal and mental demands of the task to make each scenario more dynamic and real.

2. An additional look at arrival time, by providing a reward for arriving early would also be interesting for user strategy.

3. The computer-select mode with multiple route options for the users to select initially, but operations remain the same. This would leverage the best and

minimize the least preferred parts of the app as described in this experiment.

4. A broader population to include women and non-military to improve the ability to generalize these results.

5. Customizable user options to improve initial system trust and thus performance. In the context of this experiment, users could be provided with a preset route planning options that users could be trained on and then the scenarios are designed where a particular option is more applicable. Another option would be a follow-on study where the user selects a baseline of initial route planning parameters and deals with the computer-select functionality would be informative.

6. The impact of group/team coordination on the interaction with handheld assisting devices. As mentioned in the introduction, military personnel and first responders rarely operate solo, and thus looking at the interaction of four person teams on a mobile task would be more applicable. This increased realism could manifest itself as a first responder, patrol or paint-ball task.

Appendix A

Trial Selection

Below are Tables A.1, A.2, and A.3 that define the inputs and important characteristics of each trial.

Table A.1: Trial Selections Part 1

Trial	Option	Floor Base	Stair Base	Allowed Time	Break Time
1	1	10	15	180	20
1	2	10	15	180	20
1	3	20	15	180	50
2	1	10	15	180	20
2	2	10	15	180	20
2	3	10	50	180	20
3	1	10	15	180	20
3	2	10	15	180	20
3	3	20	50	180	50
4	1	10	15	180	20
4	2	20	30	180	20
4	3	20	50	180	50
5	1	10	15	180	20
5	2	10	15	180	20
5	3	10	30	180	50
6	1	10	15	180	20
6	2	10	15	180	20
6	3	20	30	180	20
7	1	10	15	180	20
7	2	10	30	180	50
7	3	10	15	180	20
8	1	10	15	180	20
8	2	10	15	180	20
8	3	20	30	180	50

Table A.2: Trial Selections Part 2

Trial	Option	QR Time	Floor Time	Stair Time	Window	Path Nodes
1	1	4	3	10	120	56
1	2	3	2	15	120	62
1	3	4	2	10	120	68
2	1	4	3	10	120	54
2	2	3	2	10	120	68
2	3	4	2	10	120	68
3	1	4	3	10	120	51
3	2	3	2	10	120	61
3	3	3	2	10	120	73
4	1	4	3	10	120	54
4	2	3	2	10	120	66
4	3	3	2	15	120	56
5	1	4	3	10	120	54
5	2	3	2	10	120	70
5	3	3	2	15	120	70
6	1	4	3	10	120	49
6	2	3	2	10	120	61
6	3	4	3	10	120	53
7	1	4	3	10	120	54
7	2	3	2	10	150	74
7	3	4	2	10	120	76
8	1	4	3	10	120	54
8	2	4	2	15	120	56
8	3	3	2	10	120	76

Table A.3: Trial Selections Part 3

Trial	Option	Time Length	Points	QR Code	Goal List Length
1	1	182	19	3	6
1	2	178	22	4	9
1	3	174	21	4	6
2	1	183	15	3	13
2	2	179	23	5	13
2	3	182	18	4	9
3	1	181	19	3	5
3	2	172	28	4	6
3	3	181	24	5	5
4	1	183	16	3	7
4	2	175	28	5	6
4	3	179	20	4	7
5	1	183	17	3	7
5	2	182	22	4	7
5	3	180	21	3	7
6	1	182	16	3	7
6	2	180	21	4	7
6	3	181	19	4	10
7	1	183	15	3	7
7	2	175	28	5	7
7	3	182	22	4	7
8	1	183	16	3	6
8	2	181	21	3	6
8	3	178	20	4	8

Appendix B

Scenario Maps

Below are the scenario maps utilized for the six experimental trials and training trials. They are labeled with the NATO phonetic alphabet, with the first letter naming being the trial without an update and the second trial naming having the update listed in the bottom right-hand corner of the figure. For example, Alpha did not have an update and Bravo did have an update. All 28 possible QR codes are labeled on the map with their point value. The values range from one to nine. The gold star represents the goal location and the black star represents the start location.

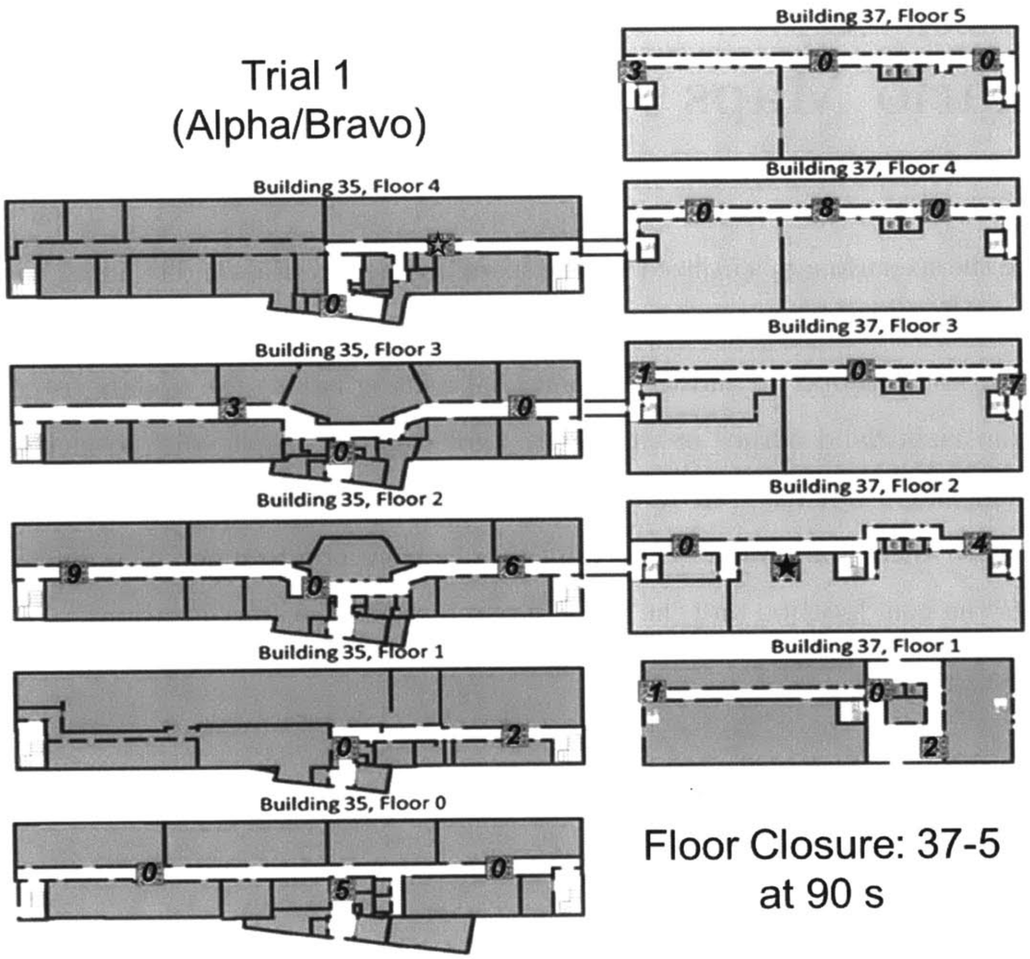


Figure B-1: Alpha/Bravo Experiment Trial 1 Scenario Map

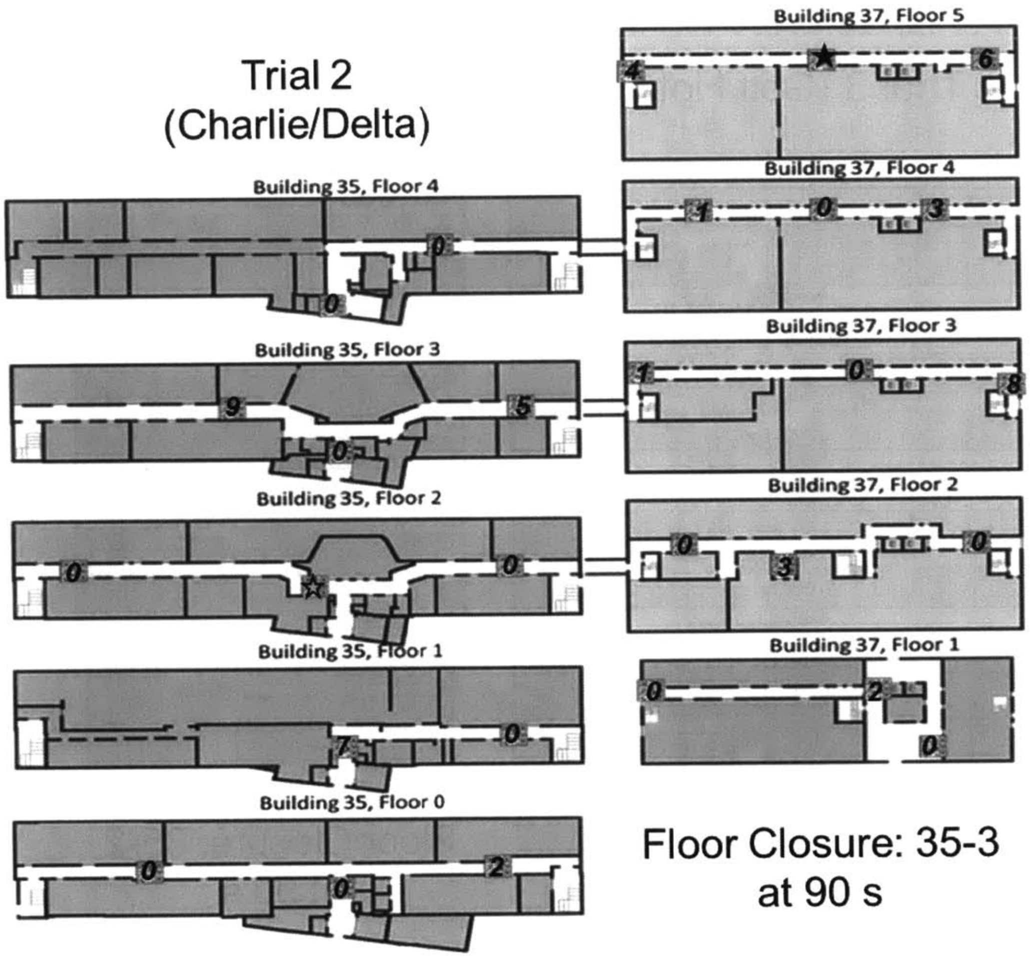


Figure B-2: Charlie/Delta Experiment Trial 2 Scenario Map

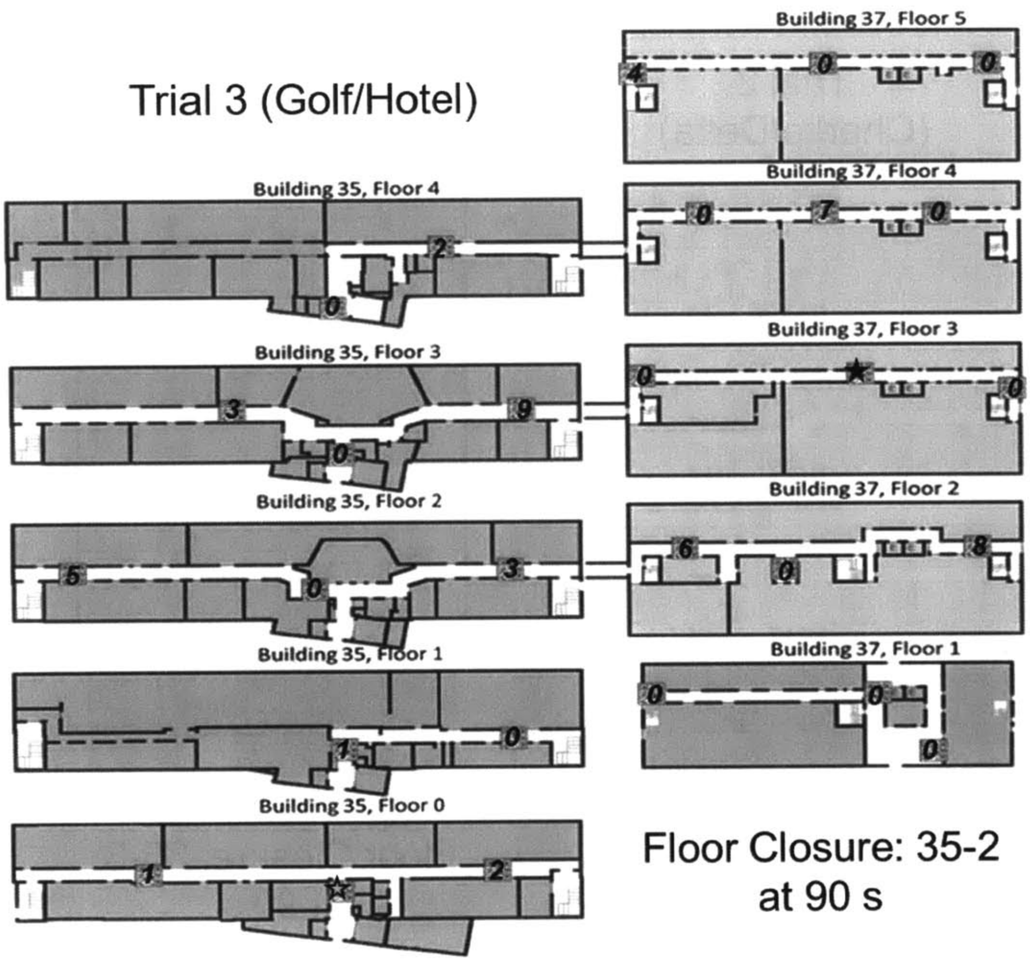


Figure B-3: Golf/Hotel Experiment Trial 3 Scenario Map

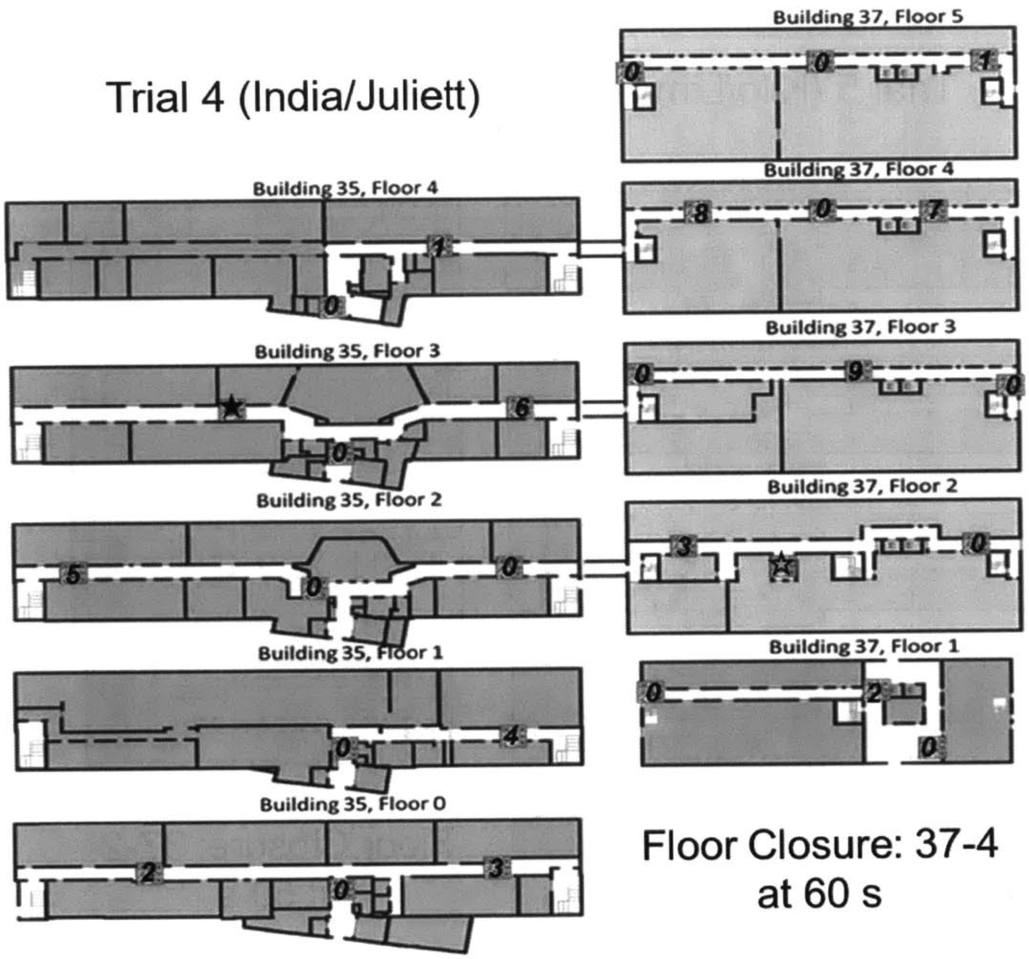


Figure B-4: India/Juliett Experiment Trial 4 Scenario Map

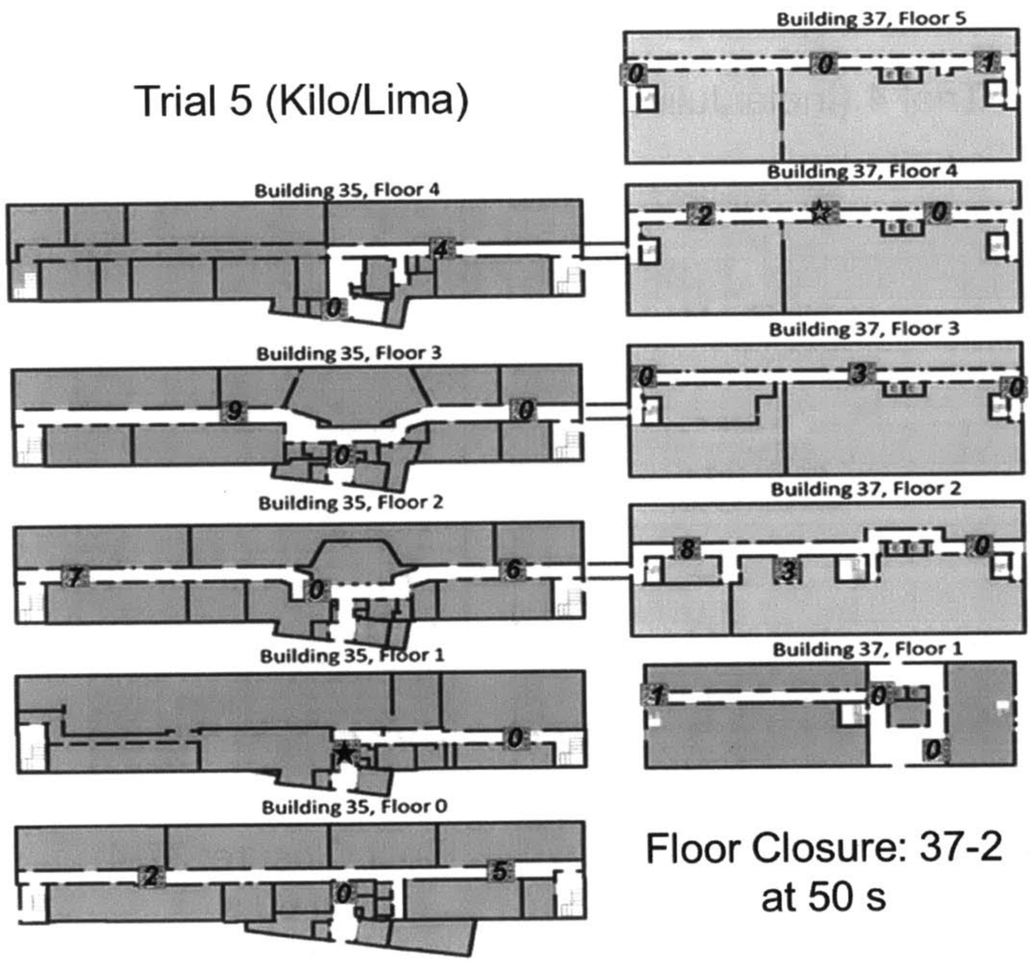


Figure B-5: Kilo/Lima Experiment Trial 5 Scenario Map

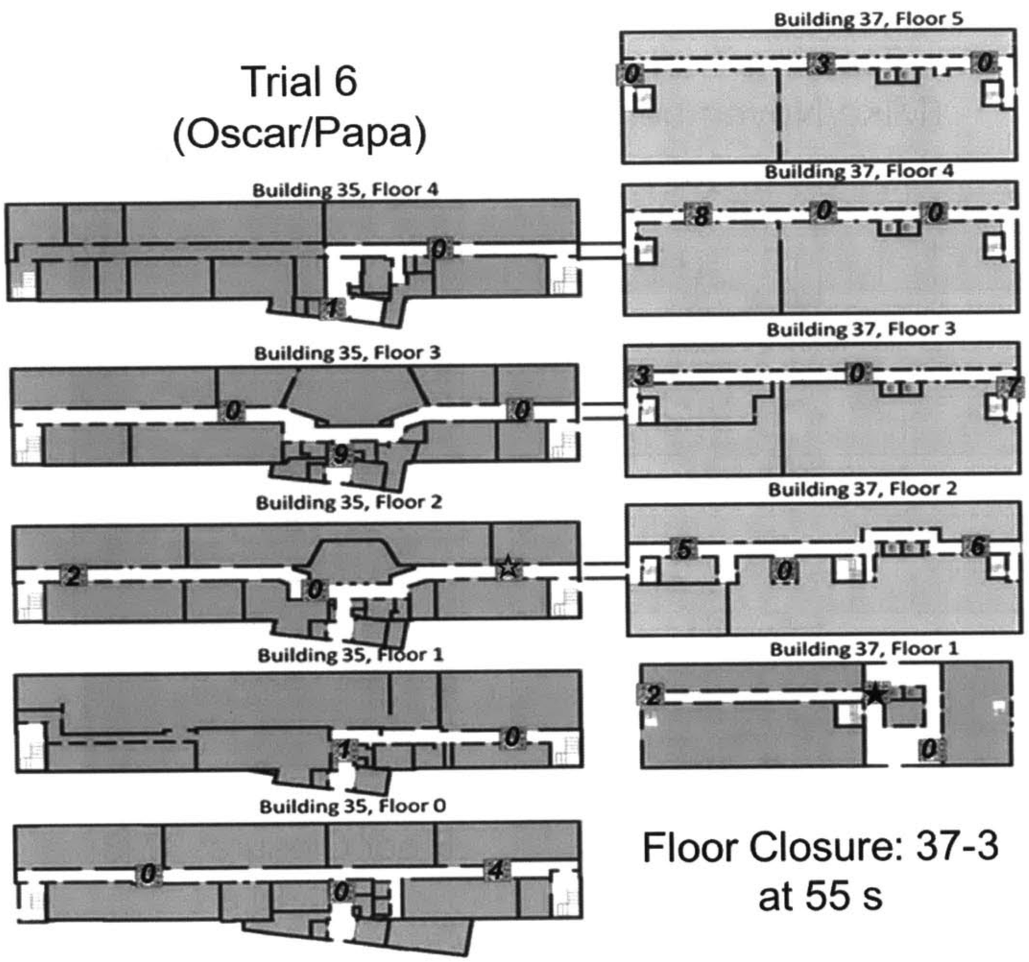


Figure B-6: Oscar/Papa Experiment Trial 6 Scenario Map

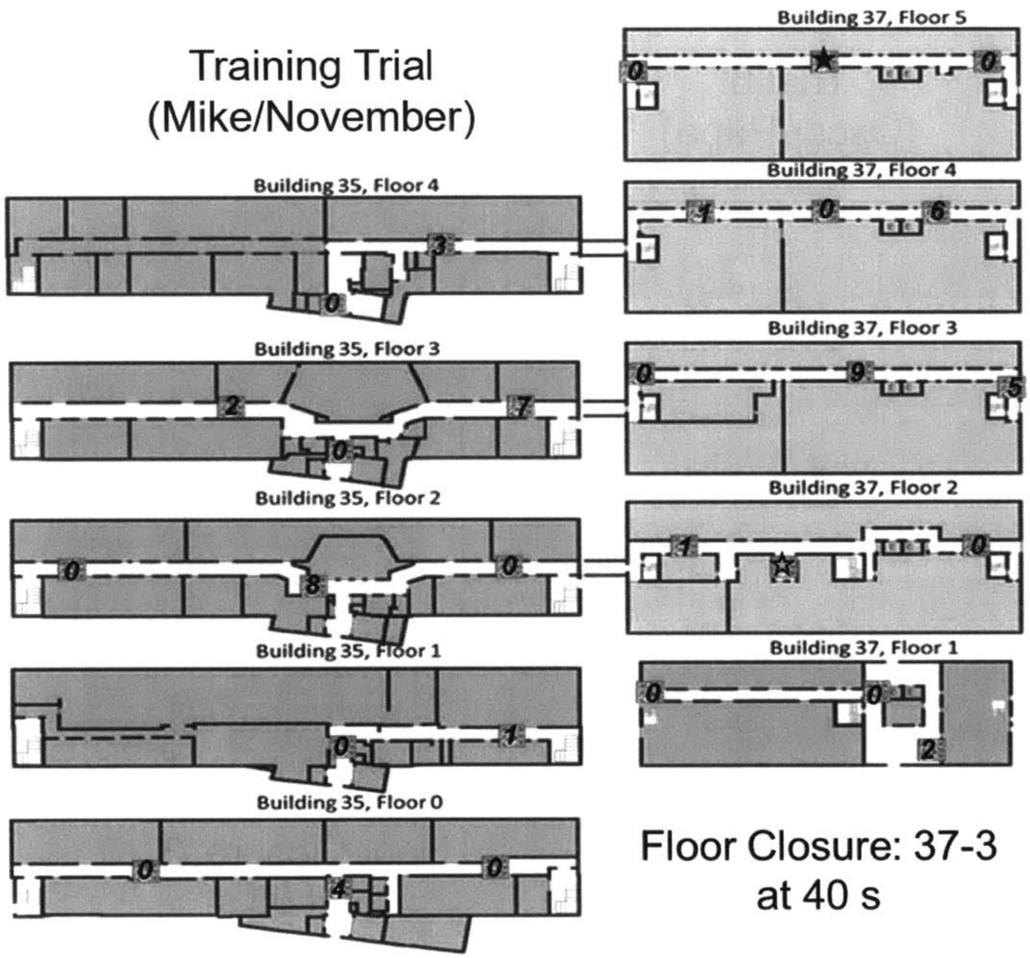


Figure B-7: Mike/November Training Trial Scenario Map

Appendix C

Scenario Maps Initial Routes

Below are the scenario maps with initial routes provided to the subjects. The red route represents the computer-select route, while the other two routes are the additional options provided by user-choice.

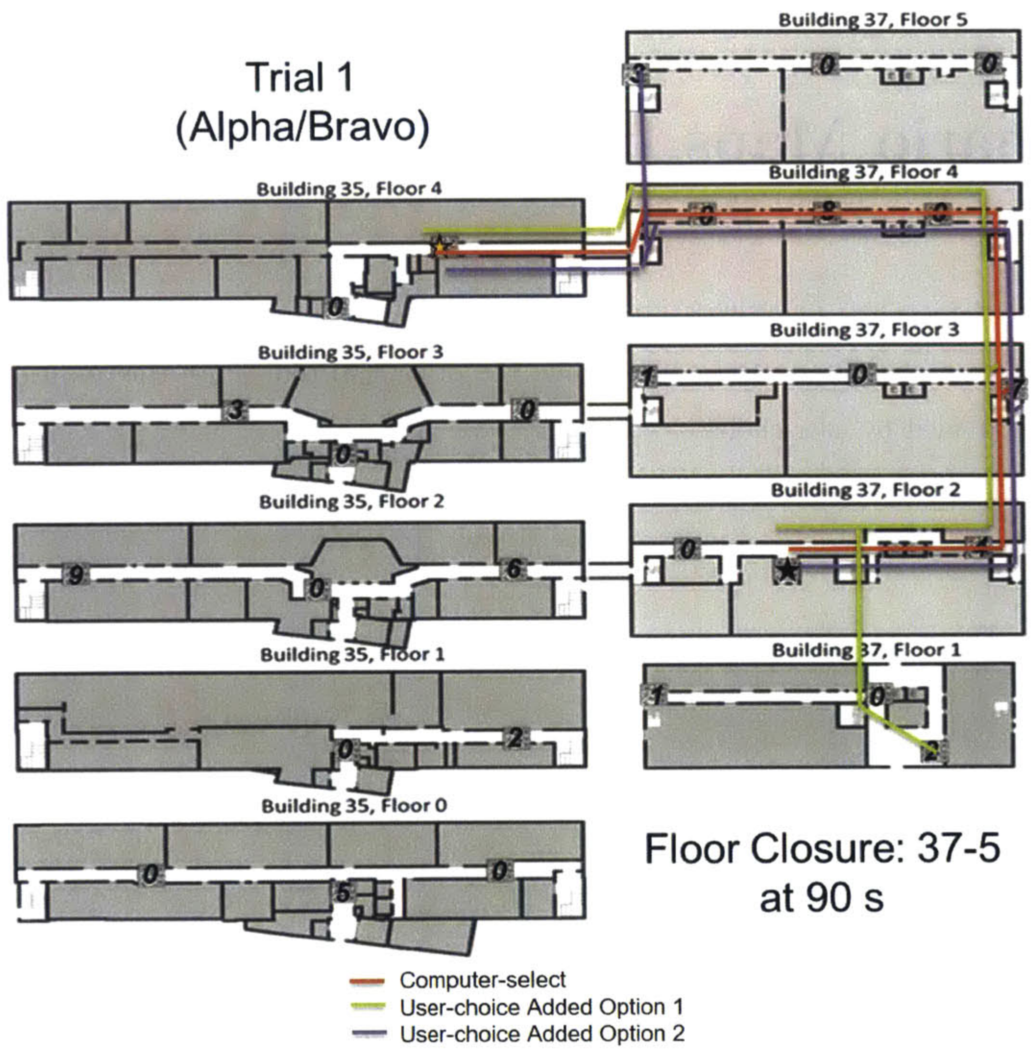


Figure C-1: Alpha/Bravo Experiment Trial 1 Scenario Map with Initial Routes

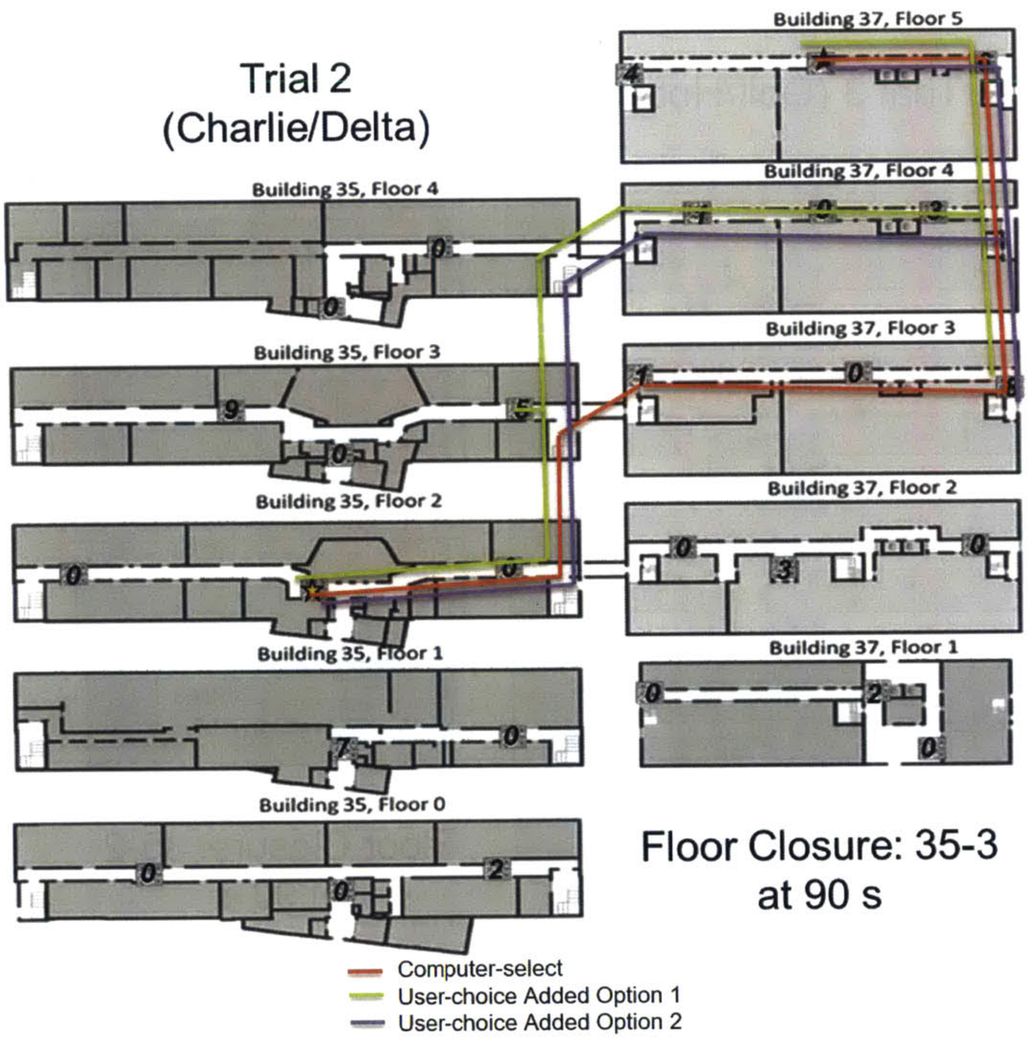


Figure C-2: Charlie/Delta Experiment Trial 2 Scenario Map with Initial Routes

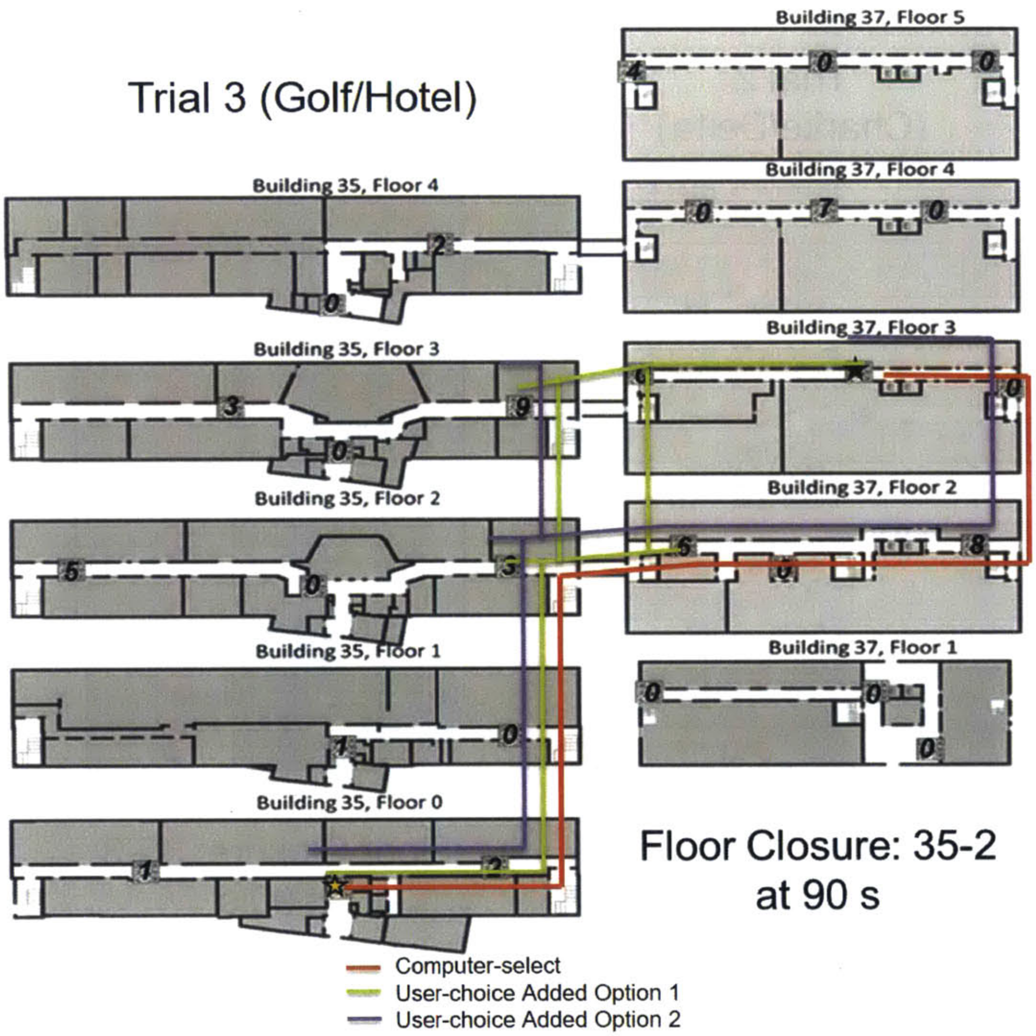


Figure C-3: Golf/Hotel Experiment Trial 3 Scenario Map with Initial Routes

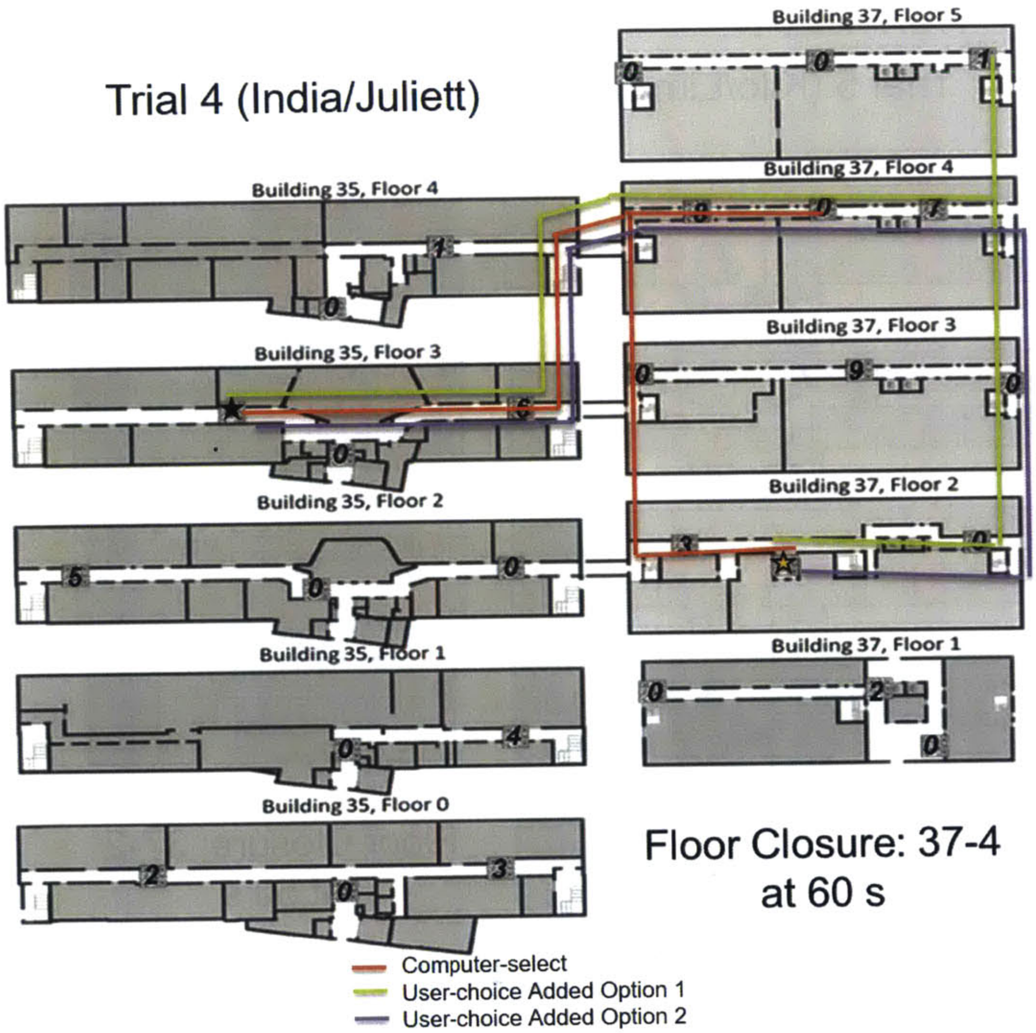


Figure C-4: India/Juliett Experiment Trial 4 Scenario Map with Initial Routes

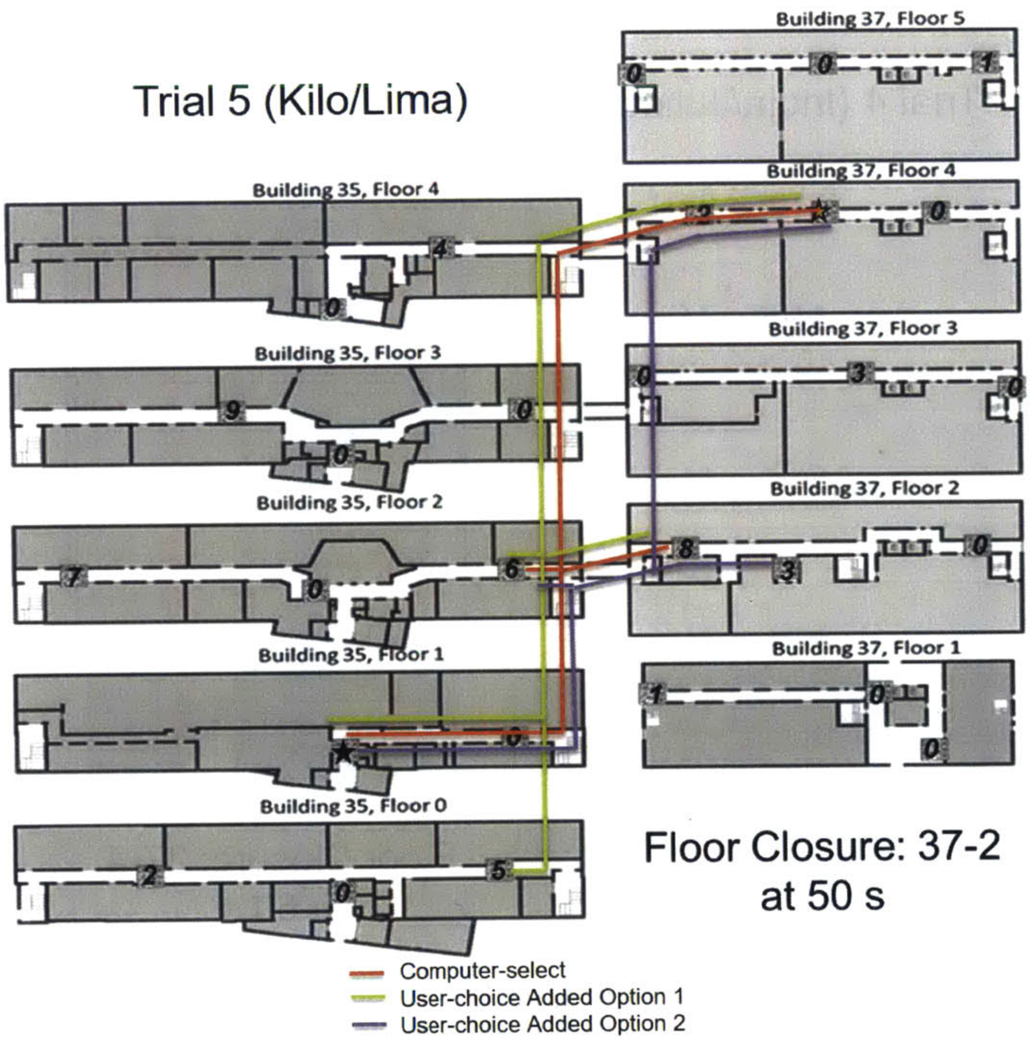


Figure C-5: Kilo/Lima Experiment Trial 5 Scenario Map with Initial Routes

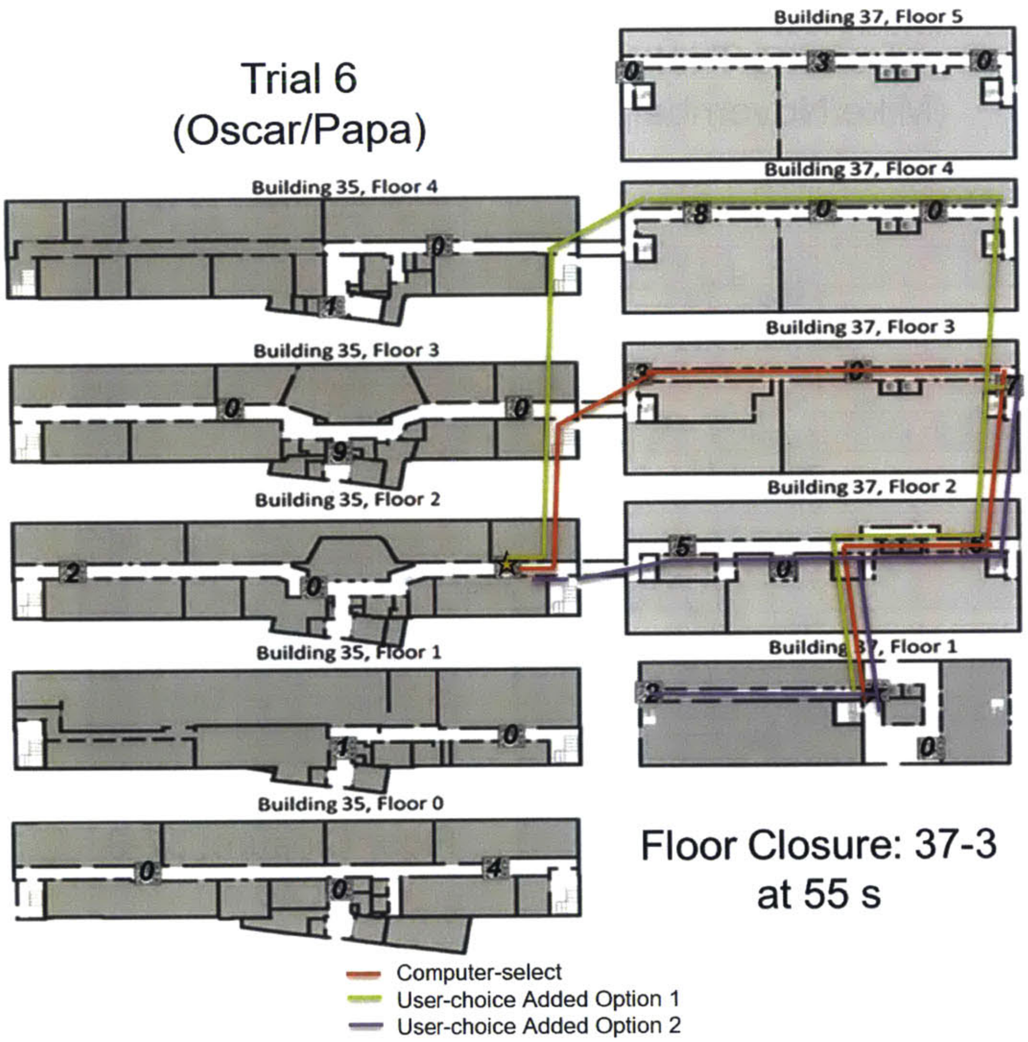


Figure C-6: Oscar/Papa Experiment Trial 6 Scenario Map with Initial Routes

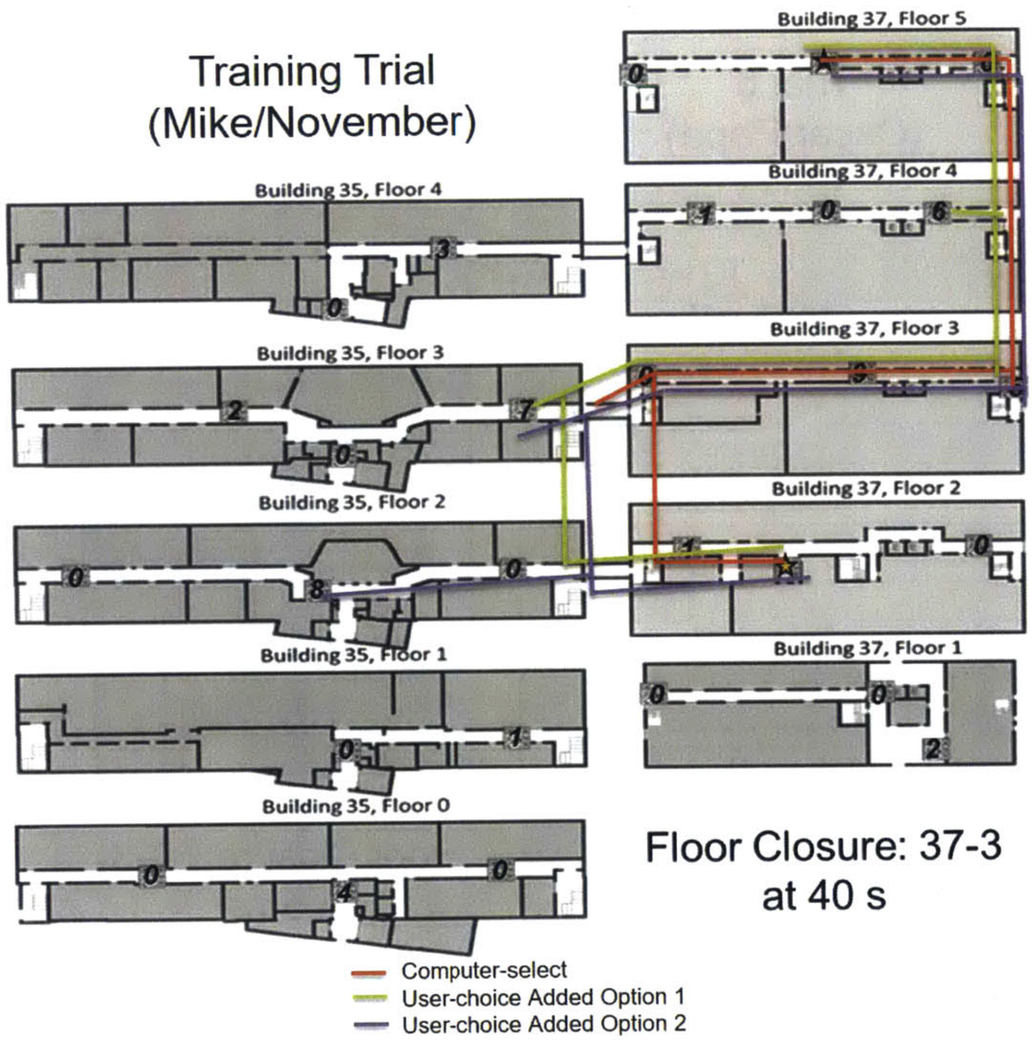


Figure C-7: Mike/November Training Trial Scenario Map with Initial Routes

Appendix D

Recruitment

Below is the recruitment email distributed to the MIT ROTC units and officer lists.

Good morning,

Have you ever wondered how handheld, automated navigation-aids affect your performance? MIT's Manned Vehicle Lab is seeking subjects affiliated with the armed forces or ROTC units to participate in a tactical search and navigation task to explore this interaction! Over three days of approximately an hour and a half each day, volunteers will be asked to navigate though the hallways of MIT and reach a specific end point in approximately three minutes while reaching as many search objects as possible. You will be compensated $15/day for participating. If you are interested please contact Tony Broll to schedule an appointment.

Thank you,

Tony Broll

awbroll@mit.edu

Appendix E

Screening

Below are the screening questions and criteria used to determine subject eligibility for the experiment. For this pilot study, all subjects were male to limit variability.

This survey will be completed in person or over the phone.

Screening Questions

Age: ________

Gender: Male or Female (circle choice)

Are you or do you suspect you may be pregnant? Y / N (circle choice)

Is your vision correctable to 20/20? Y / N (circle choice)

Are you a currently in a military training program, active duty or reserves? Y / N (circle choice)

Result of your last PRT: Pass or Fail (circle choice)

To participate subject must meet the following requirements:

Age: 18-34

Vision Correctable to 20/20

Currently in a military training program, active duty or reserves.

Passed there last physical readiness test or service equivalent.

Cannot be or be suspected of being pregnant.

For this pilot study, the subject must be male to limit variability and introducing an unknown confound in the anticipated sample.

Appendix F

Experiment Script

Below is the script used by the experimenters to ensure consistency between subject's experiences. This includes subject order and rotation of the SA characters.

Training Day:

Good morning/afternoon/evening,

Thank you for agreeing to participate in this experiment. Before we begin, I invite you to review/read this consent form. Please take your time and if there are any questions, these direct them towards me. When you are comfortable with the experiment as outlined in the consent form, please sign the last page. You are under no obligation to sign the document and signing the document does not prevent you from terminating the experiment at any time.

Pause

I'm glad you have agreed to proceed. We will begin by having you complete the following Pre-training survey on the app which will be followed by an eye exam and the perspective taking ability test (PTA test). For data storage purposes, you will be labelled as subject ____.

Pause

Now that we have completed those surveys, we will move onto familiarizing yourself with the actual experiment. Here is a training document for you to review. If you have any questions, please feel free to ask them at anytime. *Simulate the three noises following the training. Let the subject see the types of questions that are going to be asked. Log the sleep on the first printed trial question sheet.*

Pause

Now that you have completed reading the document, I'd like to ask you the questions on the last page. *Review questions and explain answers if incorrect.* Now that that is completed let's proceed with an active walk through of the buildings.

Pause

You've just had a chance to walk-through buildings 35 and 37 to get a feel for the locations of the QR codes and the stair access points. Now you'll get to do four practices trials (Mike/November) that span each condition of the study. I'll be following at a slight distance in case any issue arises.

Pause

Now that your trials are complete, I'd like to ask you those review questions again to see if it has sunk in. *Review questions again.* Thank you, now you will complete the Post training survey and then we will coordinate your following sessions.

Pause

You will meet on ______ at _____ and on _____ at ______. I recommend wearing athletic clothes. Each testing day will be roughly 2 hours.
Do you have any last questions before you go? *No.* Thank you for coming today and we'll see you on __________.

Experimentation Day:
Good _______, and welcome back. Please review the training document if you'd like and then we'll jump into the trials. Remember your focus is on 1. Arriving in time, 2. Collecting points and 3. Paying attention to SA items. In addition, can you please inform me how much you slept the past night. *Log this on the first trial question sheet.*

Do you have any questions? Otherwise we will jump right into the trials.

Conduct Trials:
Questions every trial. NASA TLX every 3rd. SA items every trial.
Change SA items every three trials.

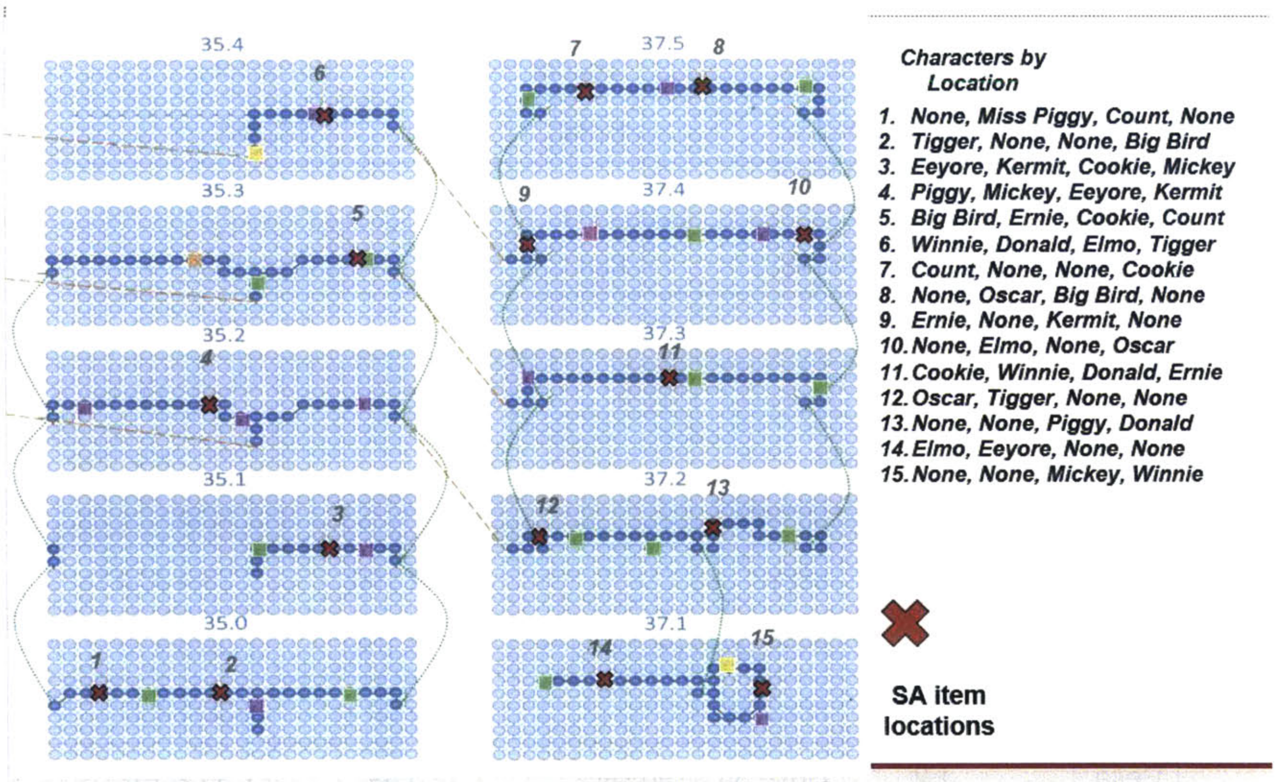
Subject	Block#	Trial#	Start
1	Computer-select	Lima	35:1:109
1	Computer-select	Golf	37:3:61
1	Computer-select	Bravo	37:2:106
1	Computer-select	Papa	37:1:63
1	Computer-select	India	35:3:105
1	Computer-select	Charlie	37:5:59
1	User-choice	Alpha	37:2:106
1	User-choice	Delta	37:5:59
1	User-choice	Golf	37:3:61
1	User-choice	Juliett	35:3:105
1	User-choice	Kilo	35:1:109
1	User-choice	Papa	37:1:63
1	User-choice	Oscar	37:1:63

1	User-choice	Lima	35:1:109
1	User-choice	India	35:3:105
1	User-choice	Hotel	37:3:61
1	User-choice	Charlie	37:5:59
1	User-choice	Bravo	37:2:106
1	Computer-select	Delta	37:5:59
1	Computer-select	Juliett	35:3:105
1	Computer-select	Oscar	37:1:63
1	Computer-select	Alpha	37:2:106
1	Computer-select	Hotel	37:3:61
1	Computer-select	Kilo	35:1:109
Subject	**Block#**	**Trial#**	**Update**
2	User-choice	Papa	37:1:63
2	User-choice	Kilo	35:1:109
2	User-choice	Juliett	35:3:105
2	User-choice	Golf	37:3:61
2	User-choice	Delta	37:5:59
2	User-choice	Alpha	37:2:106
2	Computer-select	Charlie	37:5:59
2	Computer-select	India	35:3:105
2	Computer-select	Papa	37:1:63
2	Computer-select	Bravo	37:2:106
2	Computer-select	Golf	37:3:61
2	Computer-select	Lima	35:1:109
2	Computer-select	Kilo	35:1:109
2	Computer-select	Hotel	37:3:61

2	Computer-select	Alpha	37:2:106
2	Computer-select	Oscar	37:1:63
2	Computer-select	Juliett	35:3:105
2	Computer-select	Delta	37:5:59
2	User-choice	Bravo	37:2:106
2	User-choice	Charlie	37:5:59
2	User-choice	Hotel	37:3:61
2	User-choice	India	35:3:105
2	User-choice	Lima	35:1:109
2	User-choice	Oscar	37:1:63

Name	Trial	Update	Start	Goal
Alpha	1	none	37:2:106	35:4:113
Bravo	1	close 37:5 (90 sec)	37:2:106	35:4:113
Charlie	2	none	37:5:59	35:2:132
Delta	2	close 35:3 (90 sec)	37:5:59	35:2:132
Echo	3	none	35:0:102	37:4:54
Foxtrot	3	close 35:2 (30 sec)	35:0:102	37:4:54
Golf	4	none	37:3:61	35:0:133
Hotel	4	close 35:2 (90 sec)	37:3:61	35:0:133
India	5	none	35:3:105	37:2:106
Juliett	5	close 37:4 (60 sec)	35:3:105	37:2:106
Kilo	6	none	35:1:109	37:4:61
Lima	6	close 37:2 (50 sec)	35:1:109	37:4:61
Mike	7	none	37:5:59	37:2:106

November	7	close 37:3 (40 sec)	37:5:59	37:2:106
Oscar	8	none	37:1:63	35:2:116
Papa	8	close 37:3 (55 sec)	37:1:63	35:2:116
Quebec (walk)	all 1s	none	anywhere	none



Printed Documents:

SA Survey questions and end/beginning locations
Consent forms
PTA instructions
NASA TLX instructions
Survey Examples for training day

Appendix G

Consent Form

Below is the consent form issued and signed by all eight subjects.

Approved

CONSENT TO PARTICIPATE IN NON-BIOMEDICAL RESEARCH

Trust in Adaptive and Adaptable Automation in a Tactical Search and Navigation Task

You are asked to participate in a research study conducted by L. Stirling (Ph.D. Aeronautics and Astronautics) and A. Broll (B.S. Aerospace Engineering) from the department of Aeronautics and Astronautics at the Massachusetts Institute of Technology (M.I.T.), and M. Cunha (M.Eng. Electrical Engineering and Computer Science) from the Charles Stark Draper Laboratory. This research will be in the support of A. Broll's thesis work. You have been asked to participate in this study because you are between 18-34 years of age and are affiliated with the military or an ROTC program. If you agree to take part in this study, you will be one of approximately 20 subjects. You should read the information below, and ask questions about anything you do not understand, before deciding whether or not to participate.

- **PARTICIPATION AND WITHDRAWAL**

Your participation in this study is completely voluntary and you are free to choose whether to be in it or not. Your participation or non-participation in the study will not affect any M.I.T. course grades or your relationship with M.I.T. if you are a member of the M.I.T. community. Your participation will not affect your position with the ROTC program. If you choose to be in this study, you may subsequently withdraw from it at any time without penalty or consequences of any kind. The investigator may withdraw you from this research if circumstances arise which warrant doing so. If at any time during this study, any investigator feels that your safety is at risk, the investigators may terminate your participation in this study.

- **PURPOSE OF THE STUDY**

As technology and automation continue to improve, they are implemented at all levels of the armed forces from the F-35 to the handheld devices of the infantry. As previous research has shown, the implementation of technology to resolve one issue may create an unanticipated problems. Specifically, automation has been shown to be successful at time consuming repetitive tasks, but may also cause operator confusion, increase cognitive demand, and change the user's job. Past studies have also demonstrated that the establishment of human-machine trust determines the effective and appropriate use of automation. In this study we examine how automated devices may assist the human in a dynamic environment. In this study we look at how performance and trust are affected by automation in a tactical navigation and search task. This research will help understand the interactions between operators such as first responders and military personnel in the field with handheld smart phones providing assistance.

- **PROCEDURES**

If you volunteer to participate in this study, we would ask you to do the following things:

Prior to the study, you verified via a phone conversation that the inclusion criteria have been met. This study takes place over three separate days, with one session each day. The first session is for training while the remaining two are for testing. The approximate length of the training day session will be an hour while testing day sessions last approximately an hour and a half. There will be a total of 24 trials broken in to four blocks over the two days. Each block will consist of using a navigation assist system while trying to maximize the points you collect. The paths will be located in the hallways and stairwells of Buildings 35 and 37 at MIT.

Day 1 - Training

1. At the beginning of the study, you will review the consent form with researcher. In order to continue with the experiment, you must sign the consent form. There is no obligation or pressure to sign this document. Even if signed, you are under no obligation to complete the study and may leave at any time.
2. If you sign the consent form, you will complete a survey, the Perspective Taking Ability Test and an eye test.
3. You will then receive initial training which includes an interactive demonstration with the navigation app for the experiment and a walkthrough of a hallway to demonstrate the basic procedures for the navigation assist modes. You will also learn about the search objects. You are encouraged to ask questions and test out the device.
4. Following the training, you will complete an additional survey. The second survey is used to check for clarity and understanding of the task. There will then be additional time for you to ask for clarification of the experiment. All survey questions will be digitally administered and stored.
5. At the conclusion of the survey, you will schedule your two testing day sessions. You will be advised to wear athletic gear to the testing sessions. Finally, you will be given another opportunity to ask any additional questions.

Day 2/3 - Testing

The second and third days will begin by reviewing the task and allowing you to ask questions. Trials will not be repeated if something unexpected occurs like a fire alarm. All survey questions will be digitally administered and stored. Each session will consist of 12 trials and you will be provided rest in between trials. You may wear headphones to provide you with sound cues.

Explanation of one trial:

1. You will be given a handheld mobile device with the designed application. You will be assigned one of the two automation modes and given a start and end location with a time window of three minutes to reach the desired end destination.
2. While navigating throughout the hallways and stairwells in Buildings 35 and/or 37, you will be required to scan the QR codes of the search objects in the hallways to collect points. Your goal will be to collect the maximum number of points while reaching the end location in the time allowed. The task will cover multiple floors and there will be more points available than you will have time to collect.
3. Following every trial, you will be given a short survey. You will also perform the NASA TLX, an additional survey which measures workload every third trial.
4. During random trials, you will be asked additional questions that evaluate your awareness of your surroundings.

Following your last trial of the day, there will be a final slightly longer survey.

- **POTENTIAL RISKS AND DISCOMFORTS**

Potential discomforts include fatigue due to the physical exertion of the study, tripping or falling on the stairs or in the hallways, and loud noises. To mitigate these concerns, you will be offered breaks throughout the study. In addition, the training will provide you with an opportunity to familiarize yourself with the environment. You will be penalized for speeds in excess of 4.5 mph (13.3 minute mile pace) as measured by the accelerometer in the mobile device. Any serious injury to the participant will imediately result in contact of medical services and an end to the expeiment for the subject. You may wear headphones throughout the study which may result in discomfort within your ears. You will have the ability to adjust the maximum volume of the headphones. You will have the ability to adjust the maximum volume of the headphones.

- **POTENTIAL BENEFITS**

You will not directly benefit from the study. However, we will gain a better understanding of the considerations to design a mobile automation device that could benefit first responders and military personnel.

- **PAYMENT FOR PARTICIPATION**

You will receive $15/day for participating in the experiment.

- **CONFIDENTIALITY**

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law. The results of this experiment will be identified with a participant code so your name will remain confidential. The key associating a specific code to a person will be located in a password protected folder in the lab's server. Only the key personnel will know this password. This file will be destroyed upon study completion.

- **IDENTIFICATION OF INVESTIGATORS**

If you have any questions or concerns about the research, please feel free to contact:

Principal Investigator:	Professor Leia Stirling leia@mit.edu 617-324-7410
Co-Investigator:	Meredith Cunha mcunha@draper.com
Co-Investigator:	Anthony Broll awbroll@mit.edu

- **EMERGENCY CARE AND COMPENSATION FOR INJURY**

If you feel you have suffered an injury, which may include emotional trauma, as a result of participating in this study, please contact the person in charge of the study as soon as possible.

In the event you suffer such an injury, M.I.T. may provide itself, or arrange for the provision of, emergency transport or medical treatment, including emergency treatment and follow-up care, as needed, or reimbursement for such medical services. M.I.T. does not provide any other form of compensation for injury. In any case, neither the offer to provide medical assistance, nor the actual provision of medical services shall be considered an admission of fault or acceptance of liability. Questions regarding this policy may be directed to MIT's Insurance Office, (617) 253-2823. Your insurance carrier may be billed for the cost of emergency transport or medical treatment, if such services are determined not to be directly related to your participation in this study.

- **RIGHTS OF RESEARCH SUBJECTS**

You are not waiving any legal claims, rights or remedies because of your participation in this research study. If you feel you have been treated unfairly, or you have questions regarding your rights as a research subject, you may contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T., Room E25-143B, 77 Massachusetts Ave, Cambridge, MA 02139, phone 1-617-253 6787.

Approved

SIGNATURE OF RESEARCH SUBJECT OR LEGAL REPRESENTATIVE

I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.

____
Name of Subject

____
Name of Legal Representative (if applicable)

______________________________ ______________
Signature of Subject or Legal Representative Date

SIGNATURE OF INVESTIGATOR

In my judgment the subject is voluntarily and knowingly giving informed consent and possesses the legal capacity to give informed consent to participate in this research study.

______________________________ ______________
Signature of Investigator Date

Appendix H

Simplified Survey Questions

Below is the list of questions provided to subjects to preview during training. This list is simplified and provides no information about frequency or hypothesis group.

Pre-Training Questions:

Please answer the following (All asked once):

How long have you been in the military? ______________

How frequently do you work out in a week for at least 30 minutes? _________________

How long do you do you typically work out? _________________

Which type of person would you consider yourself: Morning or Evening (circle choice)

Result of Eye Exam ____________

Result of Perspective Taking Ability Test ____________

Post-Training Questions (All asked once):

For the following statements, mark on the line your feelings for the statement from 1 (strongly disagree) to 5 (strongly agree):

I am confident in the task presented to me.

1 ————————————————————— 5

Strongly Disagree Neutral Strongly Agree

The task was clearly explained.

1 ————————————————————— 5

Strongly Disagree Neutral Strongly Agree

The task is applicable to a real scenario.

1 ————————————————————— 5

Strongly Disagree Neutral Strongly Agree

What do you expect to be the most difficult part of the experiment?

How frequently do you use GPS when driving? Walking? ___________

Which type of GPS do you use? ______________

With regards to the GPS you just referenced, mark on the line your feelings for the statement from 1 (strongly disagree) to 5 (strongly agree):

GPS route changes are appropriate.

1 ————————————————————— 5

Strongly Disagree Neutral Strongly Agree

GPS's actions are clear to me.

1 ———————————————— 5

Strongly Disagree Neutral Strongly Agree

Changes in the route displayed on the map were expected.

1 ———————————————— 5

Strongly Disagree Neutral Strongly Agree

GPS provides up to date information

1 ———————————————— 5

Strongly Disagree Neutral Strongly Agree

The reason for route changes is clear.

1 ———————————————— 5

Strongly Disagree Neutral Strongly Agree

GPS updates in the route are easily identifiable.

1 ———————————————— 5

Strongly Disagree Neutral Strongly Agree

GPS is useful for navigating.

1 ———————————————— 5

Strongly Disagree Neutral Strongly Agree

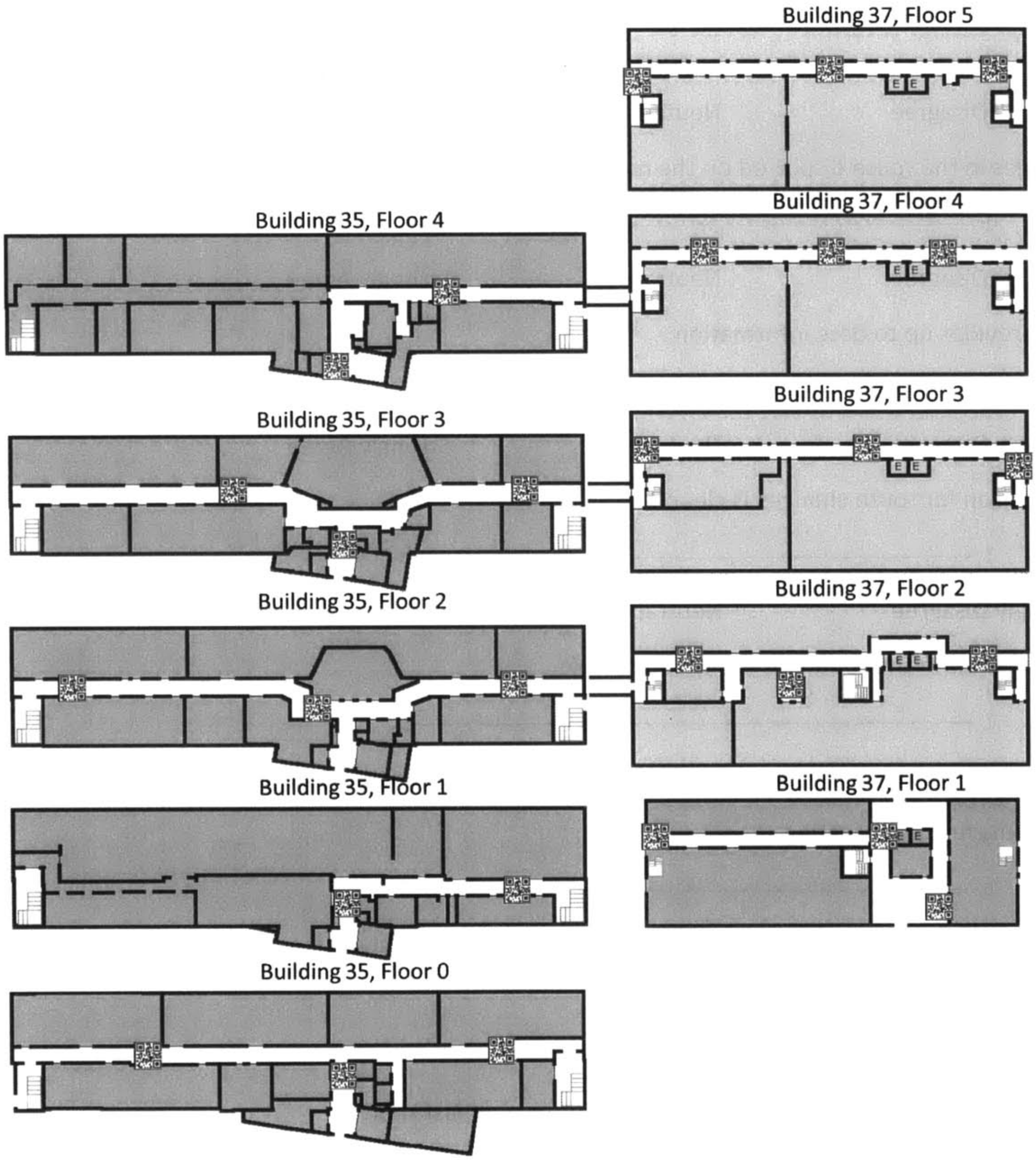
Experimental Questions:

Were there any unclaimed points available during your last trial?

If so, how many points remained unclaimed?

How many points were available on your initial route?

Mark on this map where the start and goal location were located with a 'S' for start and 'G' for goal, and the floors that you visited with an 'X':



Did you see any characters in the hall? Y / N (circle choice)

If so, where were they? Mark on the map the location with the initials of character as shown in the handout.

How much did you sleep last night? (in hours) ____________

For the block just completed, which aspect of the app was the best? The worst?

Has your strategy for the experiment changed? If so, how?

Which automation mode did you prefer? Why?

NASA TLX Index Score ____

For the following questions, mark on the line where you feel that you are for each from 1 (Low) to 5 (High):

How physically demanding was the task?

1 ———————————————— 5

Low Moderate High

How mentally demanding was the task?

1 ———————————————— 5

Low Moderate High

How much did the navigation app with the mode you just used reduce your physical demands?

1 ———————————————— 5

Low Moderate High

How much did the navigation app with the mode you just used reduce your mental demands?

1 ———————————————— 5

Low Moderate High

How fatigued are you from the task?

1 ———————————————— 5

Low Moderate High

How much do you trust the system to guide you on a path accruing the most points (including any penalties)?

1 ———————————————— 5

Low Moderate High

How complex was the task to accomplish?

1 ———————————————— 5

Low Moderate High

How applicable is the task to a real scenario?

1 ———————————————— 5

Low Moderate High

How would you evaluate your performance?

1 ——————————————————— 5

Low Moderate High

For the following statements, mark on the line your feelings for the statement from 1 (strongly disagree) to 5 (strongly agree):

The route changes were appropriate.

1 ——————————————————— 5

Strongly Disagree Neutral Strongly Agree

The app's actions were clear to me.

1 ——————————————————— 5

Strongly Disagree Neutral Strongly Agree

Changes in the route displayed on the map were expected.

1 ——————————————————— 5

Strongly Disagree Neutral Strongly Agree

The app provided up to date information

1 ——————————————————— 5

Strongly Disagree Neutral Strongly Agree

The reason for the route change was clear.

1 ——————————————————— 5

Strongly Disagree Neutral Strongly Agree

App changes in the route were easily identifiable.

1 ——————————————————— 5

Strongly Disagree Neutral Strongly Agree

The app was useful for the task.

1 ——————————————————— 5

Strongly Disagree Neutral Strongly Agree

Appendix I

Survey by Administered

Below are the survey questions organized by the frequency they were asked.

Screening Questions and Qualifications:

This survey will be completed in person or over the phone.

Screening Questions

Age: ________

Gender: Male or Female (circle choice)

Are you or do you suspect you may be pregnant? Y / N (circle choice)

Is your vision correctable to 20/20? Y / N (circle choice)

Are you a currently in a military training program, active duty or reserves? Y / N (circle choice)

Result of your last PRT: Pass or Fail (circle choice)

To participate subject must meet the following requirements:

Age: 18-34

Vision Correctable to 20/20

Currently in a military training program, active duty or reserves.

Passed there last physical readiness test or service equivalent.

Cannot be or be suspected of being pregnant.

For the pilot study, the subject must be male to limit variability and introducing an unknown confound in the anticipated sample.

All survey questions will be provided through an electronic survey on a tablet interface.

Pre-Training Questions:

Please answer the following (All asked once):

How long have you been in the military? ________________

How frequently do you work out in a week for at least 30 minutes? ____________________

How long do you do you typically work out? ___________________

Which type of person would you consider yourself: Morning or Evening (circle choice)

Result of Eye Exam ______________

Result of Perspective Taking Ability Test ______________

Post-Training Questions (All asked once):

For the following statements, mark on the line your feelings for the statement from 1 (strongly disagree) to 5 (strongly agree):

I am confident in the task presented to me.

1 ———————————————————— 5

Strongly Disagree Neutral Strongly Agree

The task was clearly explained.

1 ———————————————————— 5

Strongly Disagree Neutral Strongly Agree

The task is applicable to a real scenario.

1 ———————————————————— 5

Strongly Disagree Neutral Strongly Agree

What do you expect to be the most difficult part of the experiment?

How frequently do you use GPS when driving? Walking? ___________

Which type of GPS do you use? ______________

With regards to the GPS you just referenced, mark on the line your feelings for the statement from 1 (strongly disagree) to 5 (strongly agree):

GPS route changes are appropriate.

1 ———————————————————— 5

Strongly Disagree Neutral Strongly Agree

GPS's actions are clear to me.

1 ———————————————————— 5

Strongly Disagree Neutral Strongly Agree

Changes in the route displayed on the map were expected.

1 ———————————————————— 5

Strongly Disagree Neutral Strongly Agree

GPS provides up to date information

1 ———————————————————— 5

Strongly Disagree Neutral Strongly Agree

The reason for route changes is clear.

1 ———————————————————— 5

Strongly Disagree Neutral Strongly Agree

GPS updates in the route are easily identifiable.

1 ———————————————————— 5

Strongly Disagree Neutral Strongly Agree

GPS is useful for navigating.

1 ———————————————————— 5

Strongly Disagree Neutral Strongly Agree

Experimental Questions (Asked after every trial):

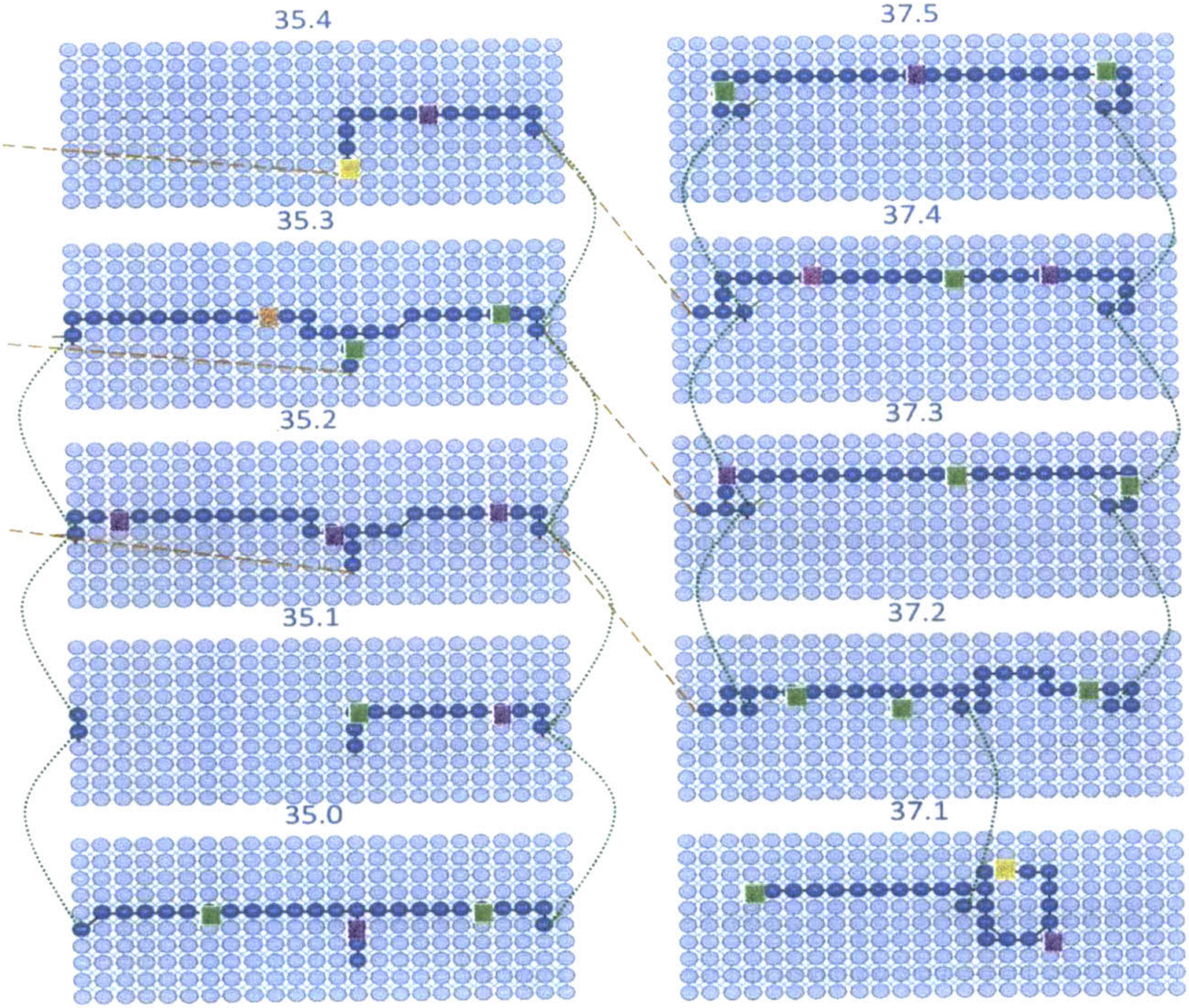
Were there any unclaimed points available during your last trial?

If so, how many points remained unclaimed?

If a road block is located on the [Enter floor number] floor, could you have claimed those points within an extra 30 seconds?

How many points were available on your initial route?

Mark on this map where the start and goal location were located with a 'S' for start and 'G' for goal, and the floors that you visited with an 'X':



For the following questions, mark on the line where you feel that you are for each from 1 (Low) to 5 (High):

How physically demanding was the task?

1 ———————————————— 5

Low Moderate High

How mentally demanding was the task?

1 ———————————————— 5

Low Moderate High

How much did the navigation app with the mode you just used reduce your physical demands?

1 ———————————————————— 5

Low Moderate High

How much did the navigation app with the mode you just used reduce your mental demands?

1 ———————————————————— 5

Low Moderate High

Has your strategy for the experiment changed? If so, how?

Experimental Questions (Asked after every 3 trials):

NASA TLX Index Score ____

For the following statements, mark on the line your feelings for the statement from 1 (strongly disagree) to 5 (strongly agree):

The route changes were appropriate.

1 ———————————————————— 5

Strongly Disagree Neutral Strongly Agree

The app's actions were clear to me.

1 ———————————————————— 5

Strongly Disagree Neutral Strongly Agree

Changes in the route displayed on the map were expected.

1 ———————————————————— 5

Strongly Disagree Neutral Strongly Agree

The app provided up to date information

1 ———————————————————— 5

Strongly Disagree Neutral Strongly Agree

The reason for the route change was clear.

1 ———————————————————— 5

Strongly Disagree Neutral Strongly Agree

App changes in the route were easily identifiable.

1 ———————————————————— 5

Strongly Disagree Neutral Strongly Agree

The app was useful for the task.

1 ———————————————————— 5

Strongly Disagree Neutral Strongly Agree

For the following questions, mark on the line where you feel that you are for each from 1 (Low) to 5 (High):

How fatigued are you from the task?

1 ———————————————————— 5

Low Moderate High

How much do you trust the system to guide you on a path accruing the most points (including any penalties)?

1 ———————————————————— 5

Low Moderate High

Experimental Questions (Asked once every 3 trials):

Did you see [insert one of the following with their picture to remind the subject which one we are talking about: Big Bird, Cookie Monster, Oscar the Grouch, Kermit the Frog, Miss Piggy, Elmo, Ernie, Winnie the Pooh, Eeyore, Tigger, Count von Count, Donald Duck, Mickey Mouse] in the hall?

If so, where was it?

Experimental Questions (Asked after every 6 trials):

For the following questions, mark on the line where you feel that you are for each from 1 (Low) to 5 (High):

How complex was the task to accomplish?

1 ———————————————————— 5

Low Moderate High

How applicable is the task to a real scenario?

1 ——————————————————— 5

Low Moderate High

How would you evaluate your performance?

1 ——————————————————— 5

Low Moderate High

For the block just completed, which aspect of the app was the best? The worst?

Experimental Questions (Asked after every 12 trials):

Which automation mode did you prefer? Why?

Experimental Questions (Asked at start of each day):

How much did you sleep last night?

Appendix J

Data Metrics by Hypothesis

Below are the data metrics grouped by principle hypotheses: background questions, trust, performance, workload and SA.

Background Questions to Reduce Confounders

How long have you been in the military? ______________

How frequently do you work out in a week for at least 30 minutes? ________________

How long do you do you typically work out? ________________

Which type of person would you consider yourself: Morning or Evening (circle choice)

Result of Eye Exam ____________

Result of Perspective Taking Ability Test ____________

For the following statements, mark on the line your feelings for the statement from 1 (strongly disagree) to 5 (strongly agree):

I am confident in the task presented to me.

The task was clearly explained.

The task is applicable to a real scenario.

What do you expect to be the most difficult part of the experiment?

How frequently do you use GPS when driving? Walking? ___________

Which type of GPS do you use? ______________

With regards to the GPS you just referenced, mark on the line your feelings for the statement from 1 (strongly disagree) to 5 (strongly agree):

GPS route changes are appropriate.

GPS's actions are clear to me.

Changes in the route displayed on the map were expected.

GPS provides up to date information

The reason for route changes is clear.

GPS updates in the route are easily identifiable.

GPS is useful for navigating.

How much did you sleep last night? (in hours) ____________

Trust Questions

For the following statements, mark on the line your feelings for the statement from 1 (strongly disagree) to 5 (strongly agree):

The route changes were appropriate.

The app's actions were clear to me.

Changes in the route displayed on the map were expected.

The app provided up to date information

The reason for the route change was clear.

App changes in the route were easily identifiable.

The app was useful for the task.

For the following questions, mark on the line where you feel that you are for each from 1 (Low) to 5 (High):

How much do you trust the system to guide you on a path accruing the most points (including any penalties)?

Performance Questions

For the block just completed, which aspect of the app was the best? The worst?

Has your strategy for the experiment changed? If so, how?

Which automation mode did you prefer? Why?

For the following questions, mark on the line where you feel that you are for each from 1 (Low) to 5 (High):

How complex was the task to accomplish?
How applicable is the task to a real scenario?
How would you evaluate your performance?

Performance Data Saved During the Trial

Points on the Computer-Select Initial Route
Points on First Route for Subject
Points Collected
Route Changes
Alerts
System Interactions
Time in Top View
Time in Side View
Time in Planning Mode
Time in QR Scan View
Time Arrived
Floors Visited

Workload Questions

NASA TLX Index Score ____

For the following questions, mark on the line where you feel that you are for each from 1 (Low) to 5 (High):

How physically demanding was the task?
How mentally demanding was the task?
How much did the navigation app with the mode you just used reduce your physical demands?
How much did the navigation app with the mode you just used reduce your mental demands?
How fatigued are you from the task?

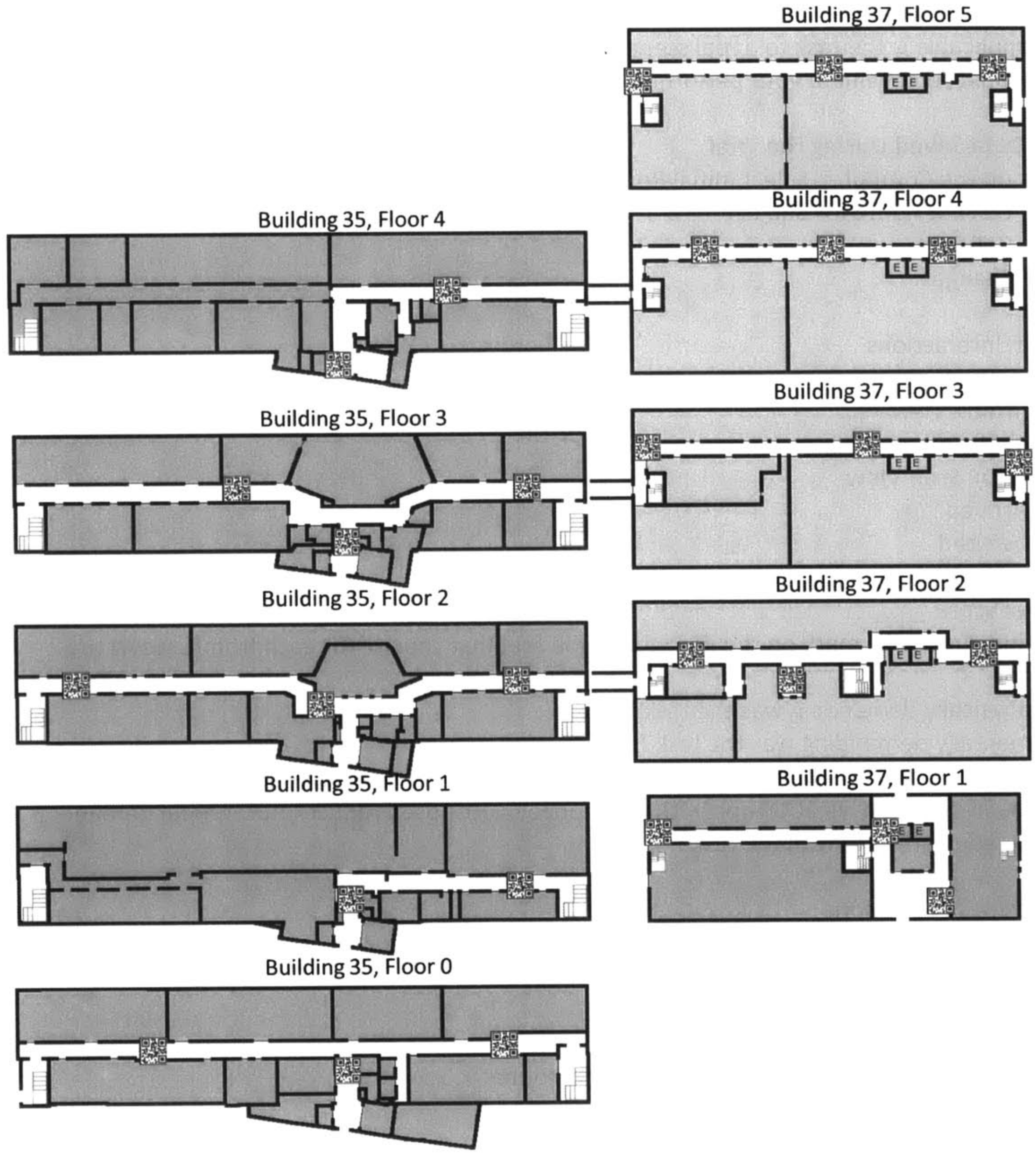
Situation Awareness Questions

Were there any unclaimed points available during your last trial?

If so, how many points remained unclaimed?

How many points were available on your initial route?

Mark on this map where the start and goal location were located with a 'S' for start and 'G' for goal, and the floors that you visited with an 'X':



Did you see any characters in the hall? Y / N (circle choice)

If so, where were they? Mark on the map the location with the initials of character as shown in the handout.

Appendix K

NASA TLX

Figure K-1 shows the sheet provided to subjects to help them complete the electronic NASA TLX. This sheet shows the first part of the survey which would then be followed by pairwise comparisons.

Figure 8.6

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Figure K-1: NASA TLX Instructions

Appendix L

Perspective Taking Ability Test

Below is the manual describing the perspective taking ability test[25].

Perspective Taking Ability Test, computerized version 1.0, by MM Virtual Design, Inc.

Thank you for acquiring our computerized Perspective Taking Ability (PTA) Test (version 1.0). Before you run this test, make sure that your computer meets the following hardware and software requirements:

- Pentium® 90MHz or higher microprocessor.
- Microsoft Windows 95 or later, or Microsoft Windows NT Workstation 4.0 (Service Pack 3 recommended) or later.
- 1024x768 or higher-resolution screen supported by Microsoft Windows (If you have a lower resolution display, please request our special version of the PTA test).
- 16 MB of RAM for Windows 95, 32 MB of RAM for Windows NT Workstation.
- A CD-ROM disc drive.
- A mouse or other suitable pointing device.
- A sound card
- Adobe Acrobat Reader (to view the Manual)

Manual

The Perspective Taking Ability (PTA) Test measures your ability to imaging different perspectives or orientations in space. This test consists of 58 trials, each containing a panel with a spatial layout of several objects. You should imaging that you are standing at one object and facing another object, and your task is to point to a third object from this facing orientation. Both the accuracy and response time are measured in this test. The typical duration time of the test is under or about 10 minutes.

Content

1. Start-up window

After you initialize the program by running the PT_Test.exe file, the following start-up panel is displayed:

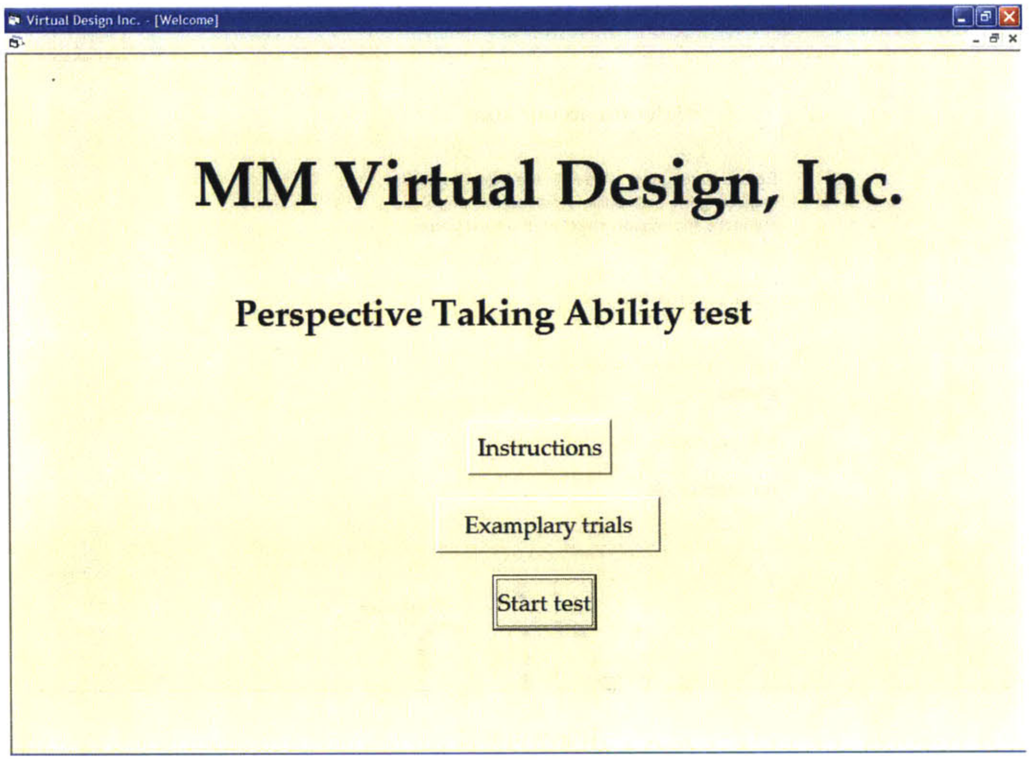


Illustration 1. The start-up window of the PTA Test.

You can choose among the following three options by clicking the corresponding button on the start-up panel:

1. Click the *Instructions* button to open the instructions.
2. Click the 3 *Sample trials* button to run three sample trials provided with the right-answer illustrations.
3. Click the *Start test* button to run the actual PTA Test.

2. Running the PTA Test

1. With the PTA Test running, click *Start test* on the start-up panel.

2. The following panel will be displayed:

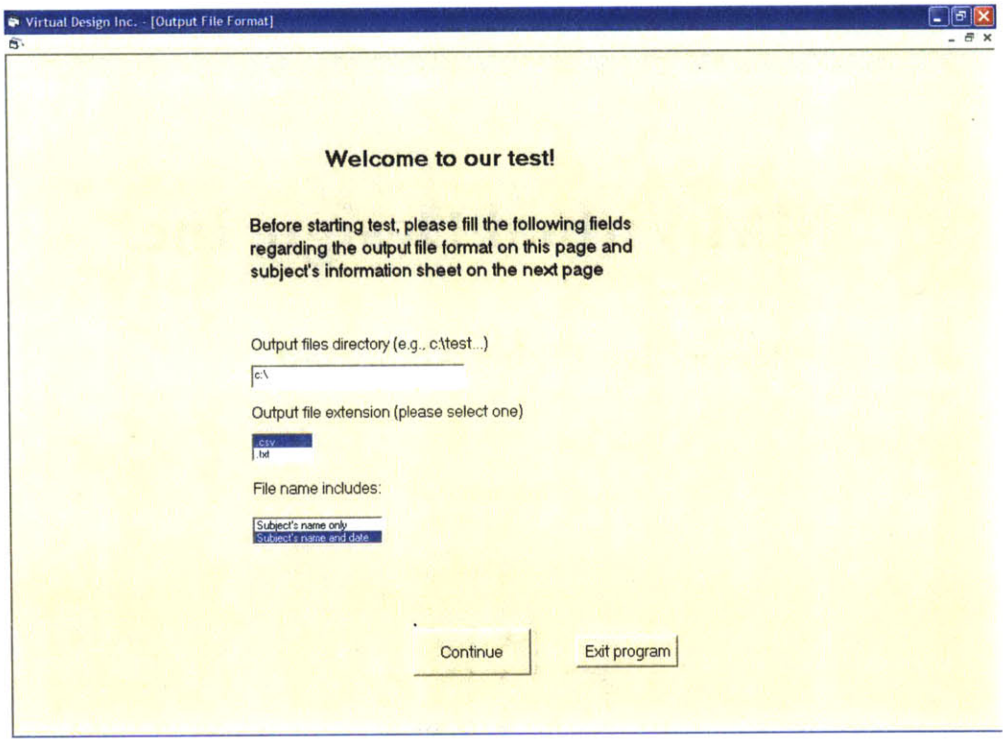


Illustration 2.

Please select

- In the first field, please type the directory name to save the test data (e.g., c:\data\PTA test). Note: you have to create this directory first.
- In the second field, please select the output file format (.txt or .csv)
- In the third field, please select the output file name format (subject's name only or subject's name + date)

Note: If you do not make a selection for the output file format, the program will prompt you to make a selection.

Click *Continue* button to proceed to the next panel

3. The following panel will be displayed:

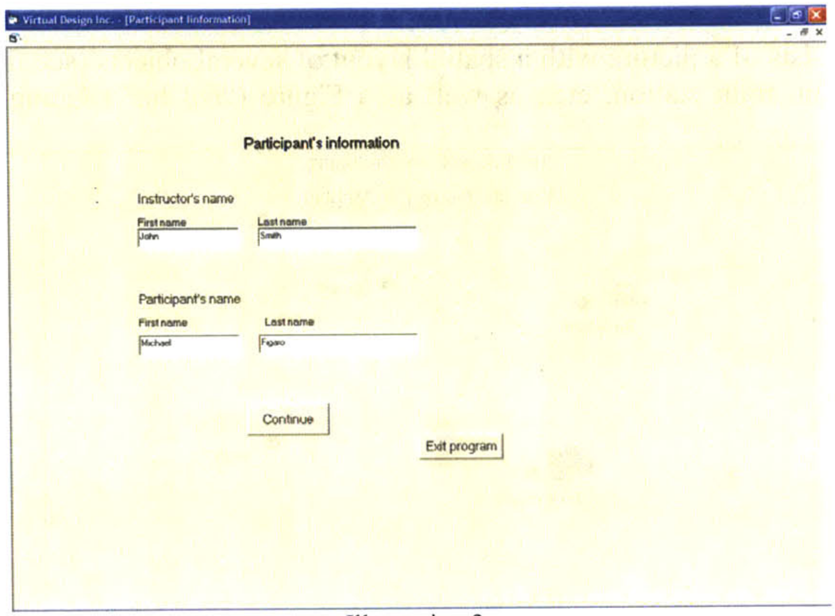


Illustration 3.

Please type in the instructor and participant names.

<u>Note</u>: If any field is left blank, the program will prompt you to fill it out.

Click the *Continue* button.

4. The following panel will be displayed:

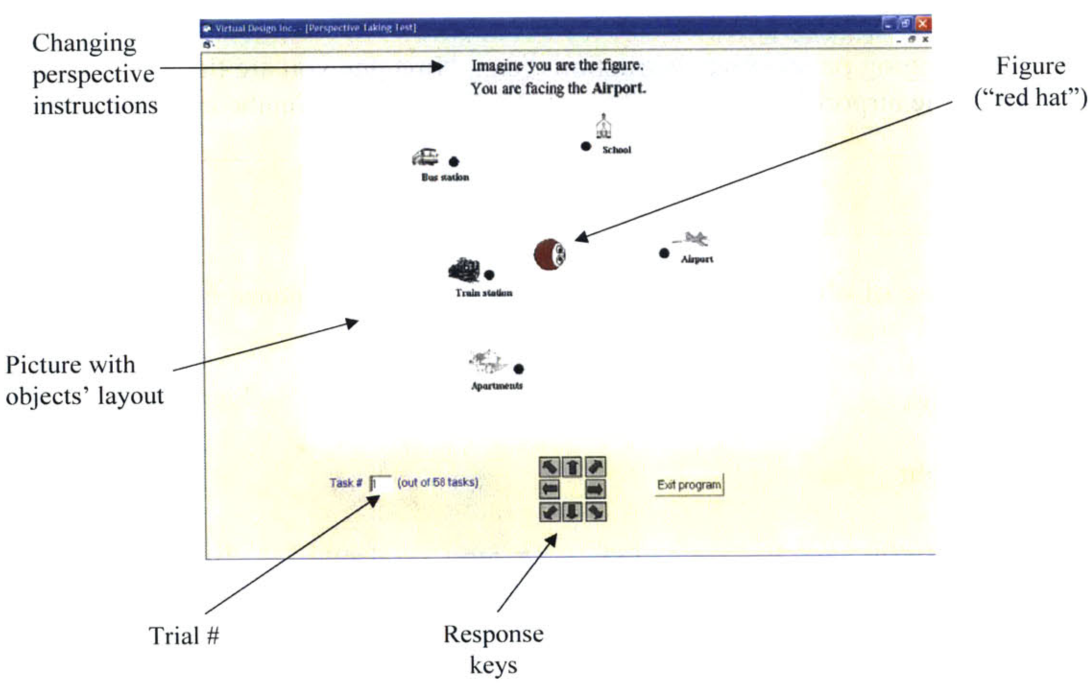


Illustration 4. Trial panel. Outside-the-panel labels are shown for explanatory purpose only and are not part of the panel.

The panel consists of a picture with a spatial layout of several objects (see figure above), e.g., bus station, train station, etc., as well as a Figure ("red hat") facing a particular object.

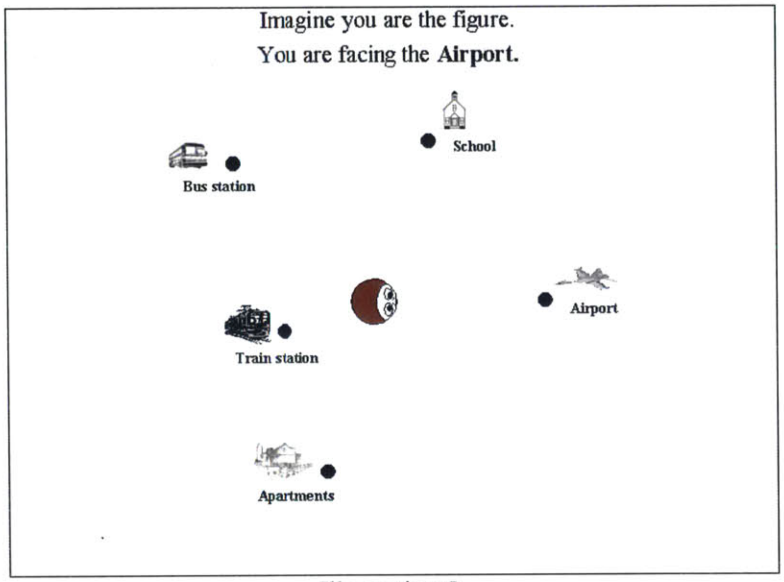


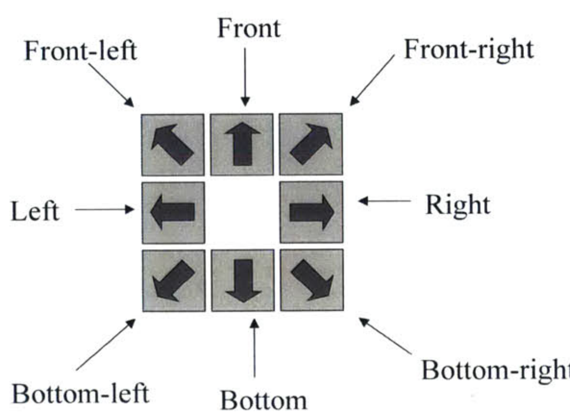
Illustration 5.

- The location of each object is given by a black circle accompanied by the object's name and schematic picture.
- The Figure ( - "red hat") represents **Your** location.
- The changing perspective instructions (e.g., "Imagine you are the Figure. You are facing the airport", as in Illustration 2 above) are shown on the top of the picture.

Imagine you are the figure. You are facing the **Airport.**

Below the picture of objects' spatial layout, there are eight response buttons – response keys.

Response options:
- Front
- Front-right
- Right
- Bottom-right
- Bottom
- Bottom-left
- Left
- Front-left



5. Imagine that **you** are the Figure ("red hat") facing a particular object (according to the changing perspective instructions). In the given example it is an airport.

6. After several seconds (~5 sec) and without you clicking any button on the panel, you will hear the "Beep" sound, and one of the objects (i.e., the corresponding black circuit), a bus station in this example, in the spatial layout will start to blink (changing the color between black and red). This is your pointing direction.

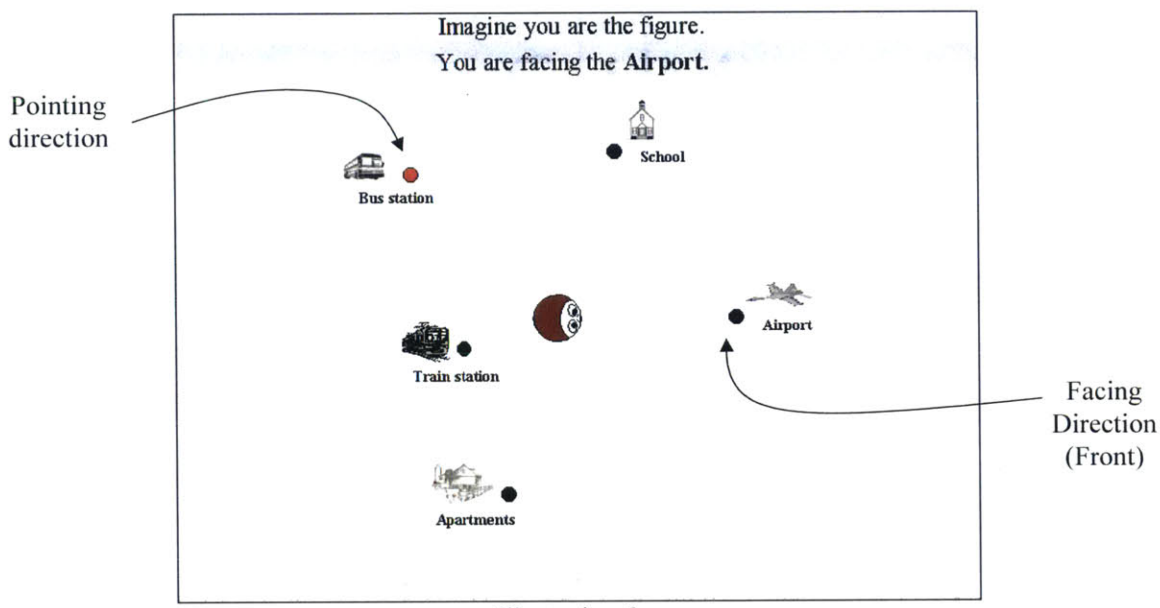


Illustration 6.

7. Point to the blinking object (a bus station in the given example) from your (the "red hat's") perspective (your facing direction is always the Front direction) by clicking the corresponding response button (Front, Front-right, Right, Bottom-right, Bottom, Bottom-left, Left or Front-left).

In the above example, the Bus Station is located to the **Bottom-left** from your (the Figure's) facing direction, so that the correct response key to click is the **Bottom-left** key:

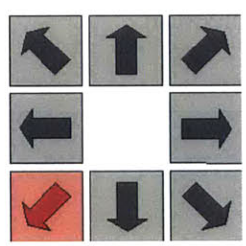


Illustration 7.

Important: In this test, your response time is also measured. Please respond as soon as possible without sacrificing the accuracy.

8. After you click any response key, the next trial panel will be automatically displayed. In the PTA Test, you will be given 58 trials. For each trial, the objects' spatial layout, the Figure position and the changing perspective instructions will change. Repeat Steps 5-7 for each trial.

It is very important that you complete all the trials in the test.

9. Upon completion of all 58 trials, the following panel will be displayed:

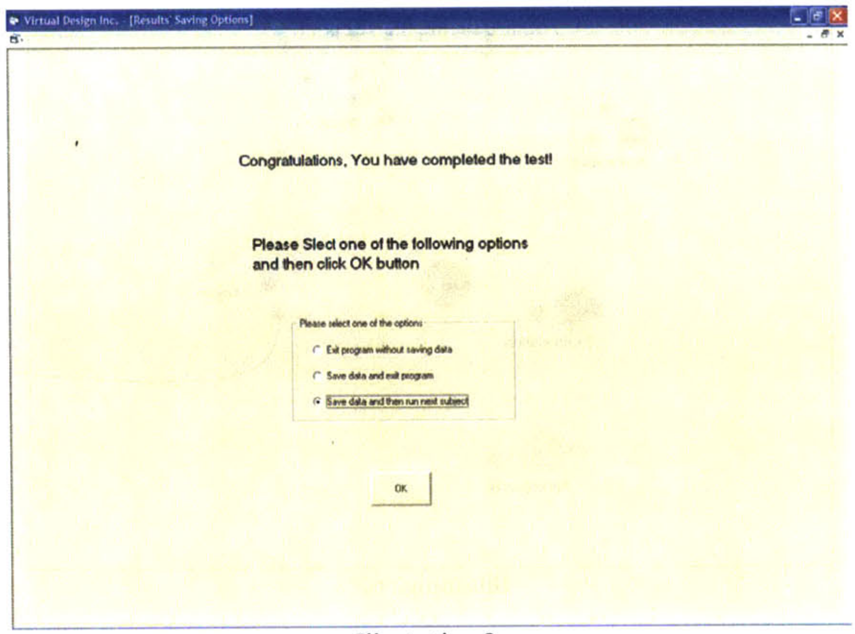


Illustration 8.

Please choose among the following three options

- Exit program without saving data
- Save data and exit program
- Save data and then run next subject

Click *OK* button to execute your selection

3. Running Sample trials

1. With the PTA Test running, click *Sample trials* on the start-up panel.

The following panel will be displayed (this panel looks the same as panels with the actual test trials, thus providing the opportunity to familiarize yourself with the actual test):

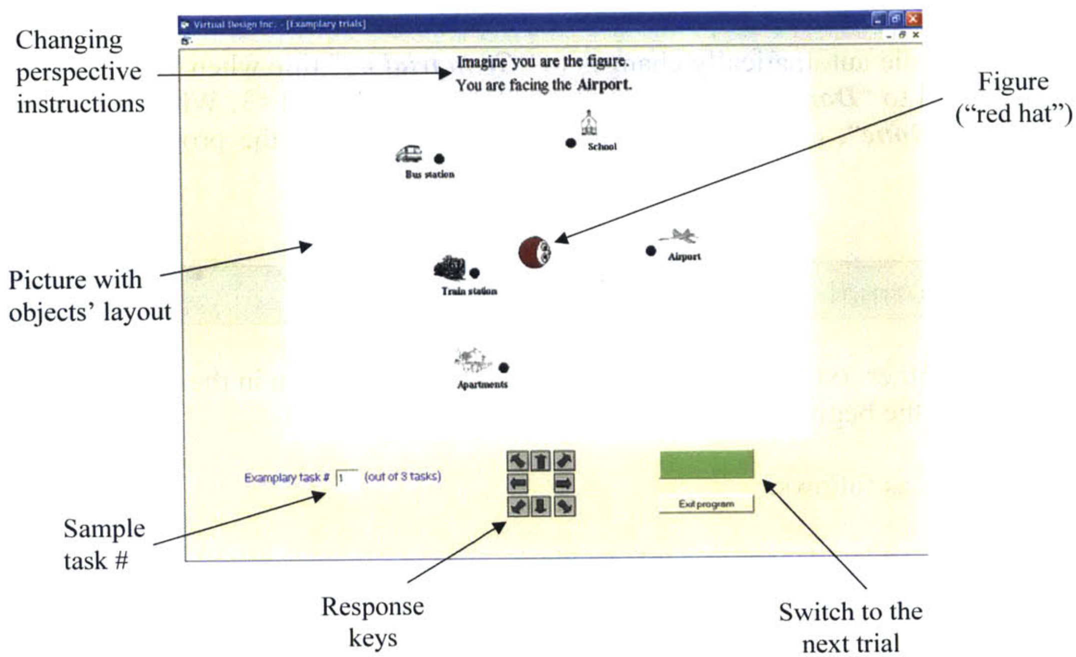
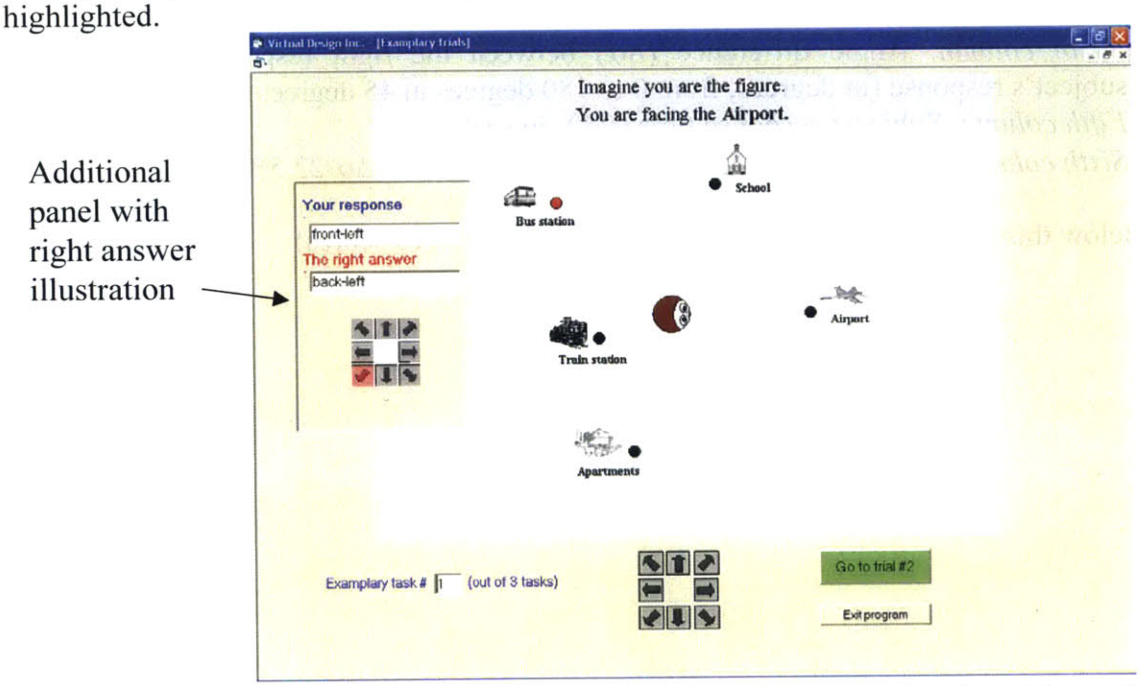


Illustration 9. Sample trials panel. Outside-the-panel labels are shown for explanatory purpose only and are not part of the panel.

Important: In contrast to the actual PTA test trials, where you are automatically switched to the next trial panel after you click any response key, for each of the sample trials, after you click any response key, an additional panel is displayed on the left side of the spatial layout panel. This panel shows two text messages with your response and with the right response, as well as the picture of the response key array with the right key highlighted.



To go to the next sample trial, click ***Go to trial #2*** button (green-colored button) to proceed to the next sample trial.

Note: This button title automatically changes to "***Go to trial #3***" title when you are doing sample trial #2 and to "***Done***" title when you are doing sample trial #3. When the button title changes to "***Done***", clicking this buttons will return you to the program start-up window.

4. Output file format

The output file is either .txt or .csv format (depends on your selection in the *Output File Format* window at the beginning of the PTA test).

The file structure is as follows:

1. First three rows display descriptive information:
   First row: Date: "m-d-yy".
   Second row: Instructor: "Instructor's name".
   Third row: Participant: "Participant name".

2. It is followed by 59 rows x 6 columns block.
   For each column, first row is a heading followed by 58 rows of the data
   First column: Trial #
   Second column: Right response key (in degrees; counting clockwise from the facing direction)
   Third column: Subject's input (in degrees; counting clockwise from the facing direction)
   Forth column: Angle difference ($\Delta\alpha$) between the right response key and the subject's response (in degrees, from 0 to 180 degrees in 45 degrees increments).
   Fifth column: Subject's response time (RT) (in sec).
   Sixth column: Itemized score (formula: $100/((RT+2)*(1+(\Delta\alpha/22.5)^2)))$)

3. Below this 59x6 block is total score (mean of 58 itemized scores)

Appendix M

Training Document

Below is the training document provided to all subjects with details on how to utilize the app.

IndoorNav Operation Instructions

Experiment Overview

This study takes place over three separate days, with one session each day. The first session is for training while the remaining two are for testing. Over the two testing days, there will be a total of 24 trials broken into four blocks. Each block will consist of using a phone-based navigation aid while trying to maximize the points you collect as you walk through the hallways. The paths will be located in the hallways and stairwells of Buildings 35 and 37 at MIT. You will not be permitted to use the elevators.

At the beginning of each trial, you will be assigned one of the two automation modes (computer-selected or user-choice). You will then be given a start and end location with a time window of three minutes to reach the desired end destination. In computer-selected mode, the app will provide all route planning functions without the input of the user. While in user-choice mode, the app will require input from the user to select one of three routes to traverse based on points available, number of staircases and estimated time.

While navigating throughout the hallways and stairwells in Buildings 35 and/or 37, you will collect points by scanning the QR codes in the hallways. Your goal will be to collect the maximum number of points while reaching the end location in the time allowed. The available navigation area spans multiple floors and there will be more points available than you will have time to collect. During some trials, you will receive updates that may change what floors are available, where your end location is on the map, or where points are located. In computer-selected mode, the app will automatically reroute you to account for the update. While in user-choice mode, the app will require you to select a new route from options that account for the update.

Completing the Trial

You have 180 seconds to navigate through the buildings, scan QR codes for points, and scan the goal QR code.

The first priority is reaching the goal QR code and scanning it before the trial time is up.

The second priority is maximizing points by scanning additional QR codes along your route.

The third priority is the secondary situation awareness task which will be explained below in the **Situation Awareness Task.**

Your score will be penalized if you don't scan the goal QR code before the trial time has run out. You will lose a fourth of your points for being up to 10 seconds late. After that, the penalty will become half of your points. Anything more than 30 seconds late will be considered an incomplete trial and you will receive no points. **There is no penalty or bonus for arriving early.**

There are also penalties for using closed-off areas within the two buildings, including floors that might become unavailable after receiving an update. A deduction of 5 points will be given for crossing a closed floor. The staircases of a closed floor are still permitted. In addition you are not permitted to use elevators, therefore the use of elevators will result in a loss of all of your points.

For your safety as well as that of faculty and students who may be in the hallways during the experiment, you are not to exceed the speed of a fast walk at any time. You will be assessed a penalty of half of your points for exceeding on average 4.5 mph which is approximately the preferred transition from walking to running.

Violation	**Penalty**
Late <10 seconds	¼ Point Total
Late 10 - 30 seconds	½ Point Total
Late >30 seconds	All Points
Cross Closed Off Area	5 Points
Average > 4.5 mph	½ Point Total
Use an elevator	All Points

The Two Navigation Modes

The two different navigation modes that you will be using separately are computer-selected and user-choice.

In **computer-selected** mode, you will be given a navigation route to maximize points and reach the goal QR code in time. The computer-selected route is dynamic and will incorporate your deviations within approximately 10 meters. For example if you want to scan a QR code that you can see on your map that is nearby, but that is not included on your route. It won't necessarily provide a new route, but will include you detour as part of your route .

If you wander from the route a significant distance, for example to a different floor than was included on the route, then the navigation system will calculate a new route for you. If a new navigation route is calculated for you, a sound notification will play and a message will appear in the Alert Bar for you to see.

In **user-choice** mode, you will be given three navigation routes to pick from to maximize points and reach the goal QR code in time. You will select a route from three options at the start of the trial and then can change the route to follow at anytime during the trial through the Alert Bar.

Similar to computer-selected, the user-choice navigation route will adjust itself if you wander from it a small distance, for example if you want to scan a QR code near to but not included on your route.

However, your navigation route is not dynamic as in the computer-selected mode. If you wander from the route a significant distance, then you will be prompted to decide whether to choose a new route or not. If you wandered too far off path, then your current route could be too long to reach the goal in time. Refusing a new shorter route recalculation could result in failure to successfully complete the trial within 180 seconds.

The Map Interface

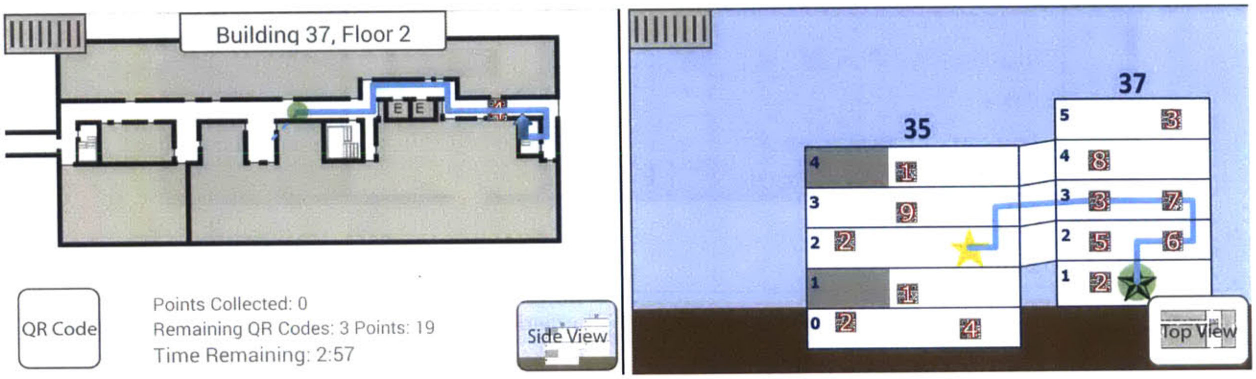


Figure 1: Top View and Side View

The maps of building 35, and 37 are displayed in the app with two different perspectives, one from above called the Top View and another from the side called the Side View. The Top View shows the floor plan of the current floor from above, while the Side View shows the global perspective of the three buildings.

Understanding the Top View

The **Top View** shows the floor plan of your current floor from above. Your current position is shown with a green circle on the map.

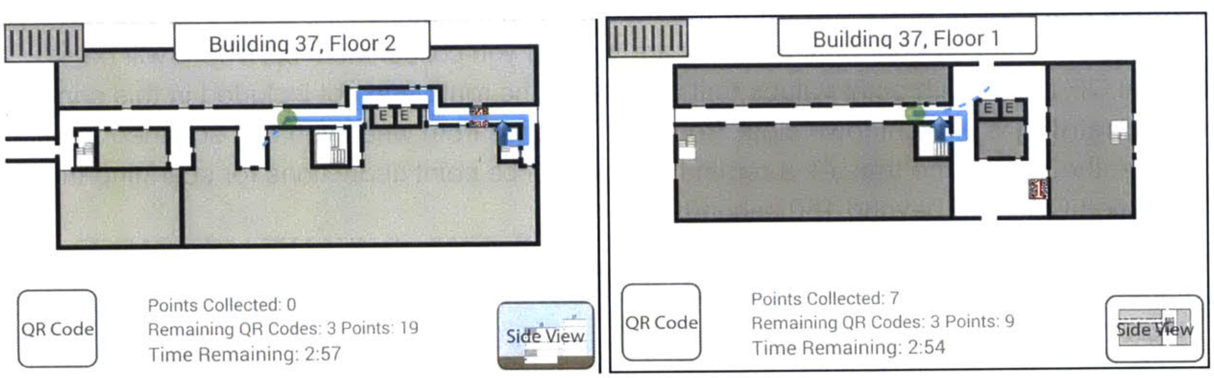


Figure 2: Top Views

The floor plan and map displays do not rotate to align North nor to align with the phone orientation, instead they are static like viewing a real map.

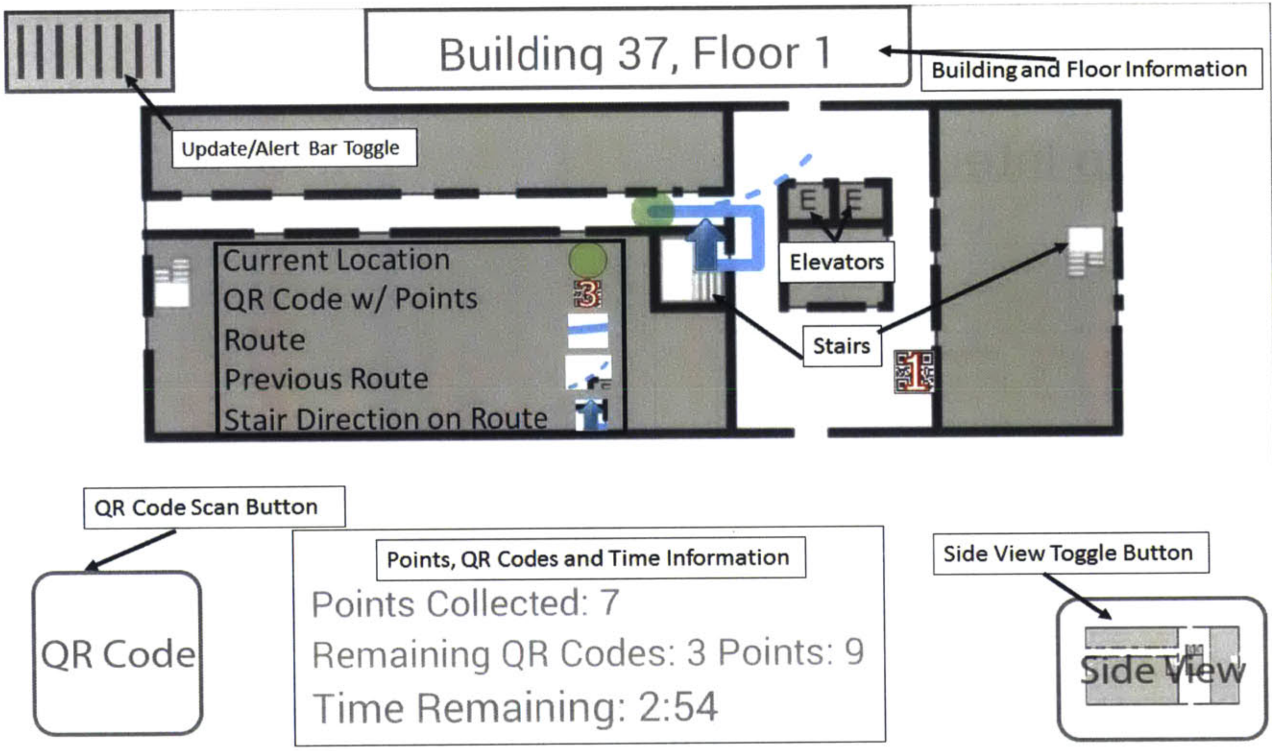


Figure 3: Labeled Top View Map

A solid blue path is drawn to show your currently-planned route. A dotted blue path is also drawn to show the route that you have already taken in this trial. The nearby QR Codes are shown with a small QR square and a red number equal to the associated point value.

Important information is organized on the the Top View display below the floor plan. The **Points Collected** shows the number of points successfully collected during this trial. The **Remaining QR Codes** and **Points** below that show the number of QR codes on the route that you have yet to scan and the total number of points they are worth if you collect them all. There will be additional QR codes with point values that are not on the route and not included in this sum. **Time Remaining** is a countdown clock from 180 seconds from when you first scanned the start location at the start of the trial. As a reminder there will be point deductions for scanning the goal location QR code beyond 180 seconds.

From the Top View, you can access the **QR Code** button and the **Side View** button.

The Side View button can be pressed anytime during the trial to toggle between the Top View and Side View displays. The Side View map shows a global perspective of the trial and is useful for understanding the route ahead. See Understanding the Side View below for more information.

Understanding the Side View

The **Side View** shows a global perspective of your trial by drawing the navigation route across the two buildings and by showing you all of the available QR Code locations and associated points for the trial. From the Side View, you can quickly reorient yourself within the trial to see, for example, how many floor changes are on the route ahead of you.

Figure 3 shows the actual positions of Buildings 35 and 37 relative to each other. In order to display both buildings side-by-side we took the view from the bottom of the map, with Building 37 on the right and Building 35 on the left to create a 2D map.

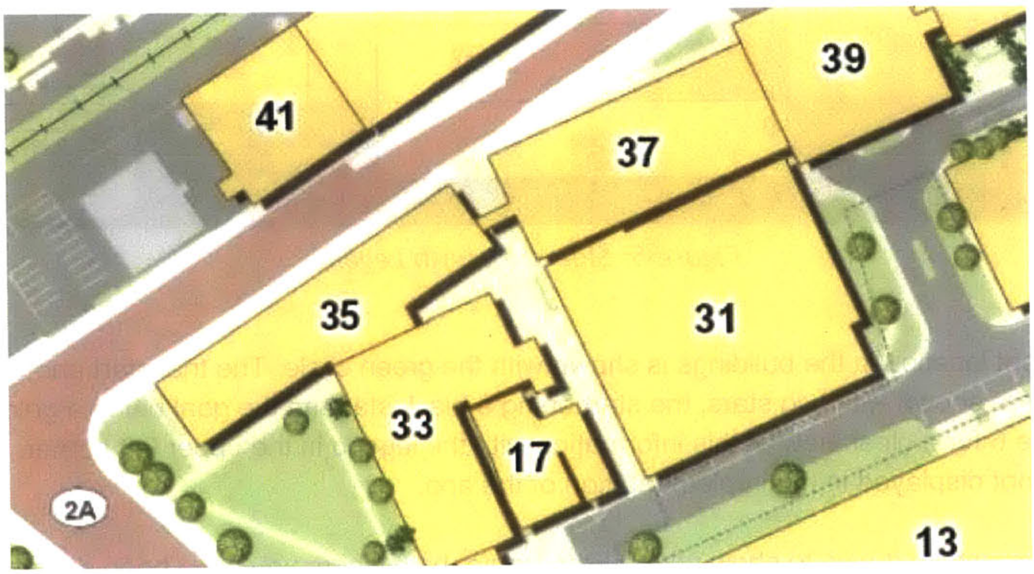


Figure 4: MIT map of Buildings 35 and 37

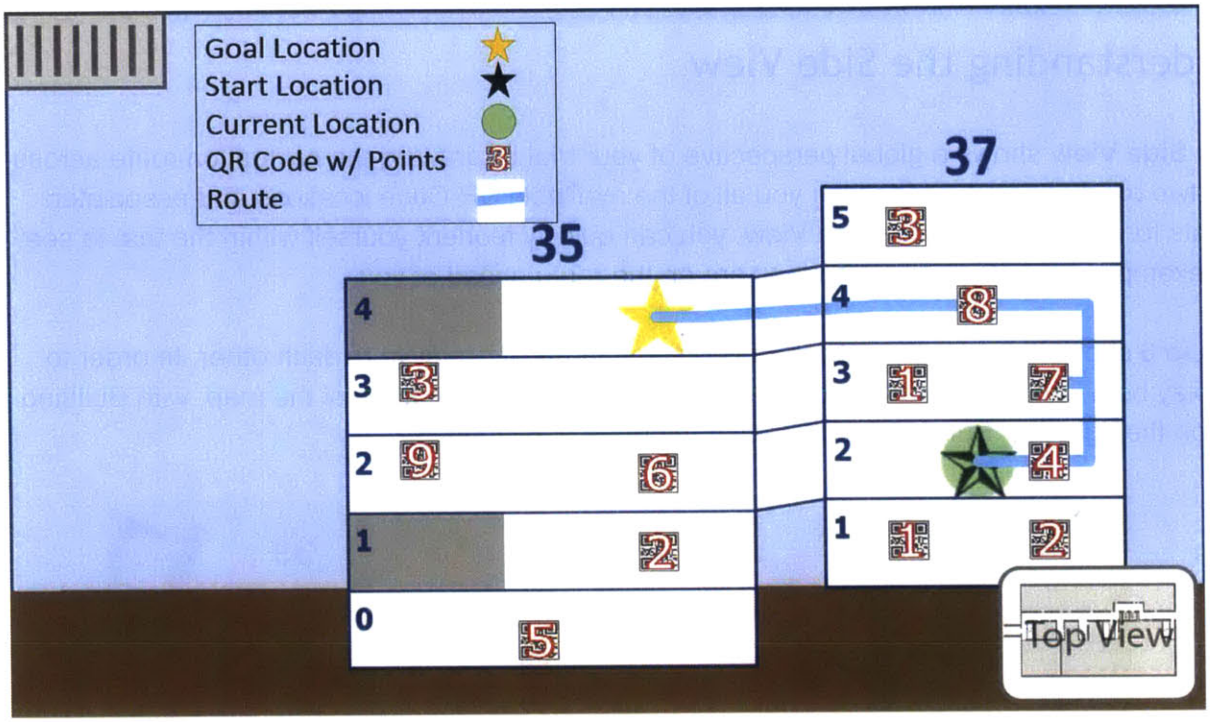


Figure 5: Side View with Legend

Your current location in the buildings is shown with the green circle. The trial start and goal locations are shown with two stars, the start being a black star and the goal being a gold star. See Figure 6 for a clear view of this information with the legend in the upper left corner. The legend is not displayed in the implementation of the app.

A solid blue path is drawn to show your current route, both where you have been so far and where you still need to go in order to reach your goal. You can see where your path might come close to a valuable QR Code and take necessary action if you have enough time to spare.

The Side View is static like the Top View in that the map does not rotate to align North nor to align with the phone orientation.

The Side View button can be pressed anytime during the trial to toggle between the Top View and Side View displays.

Toggle back to the Top View in order to scan posted QR codes

Simply press the Side View button again, when it says Top View, in order to return to the Top View display where you have access to the QR Code button and important information about the current trial.

The Alert Bar

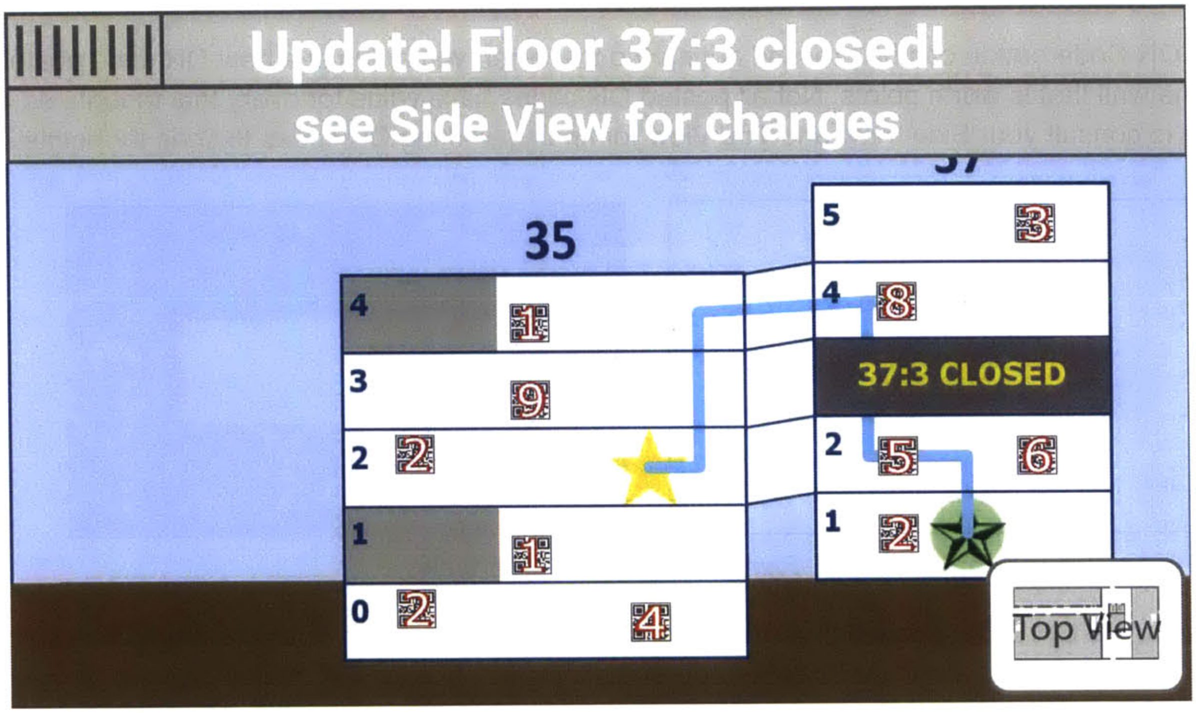


Figure 6: Alert Bar Example

The Alert Bar will inform you when an update occurs and what the specific update was. When an update occurs, the Alert Bar immediately pops up and an alert sound is heard. The Alert Bar can be toggled by pressing the button in the top left corner. The Alert Bar also allows the user to change the path in the user-choice mode either due to the update, or just in general.

The QR Code Scan View

The QR Code button can be pressed during the trial when you arrive at a new QR code posted on the wall that is worth points. Not all posted QR codes have value for every trial though, so be sure to consult your Side View and Top View for finding working QR codes to scan for points.

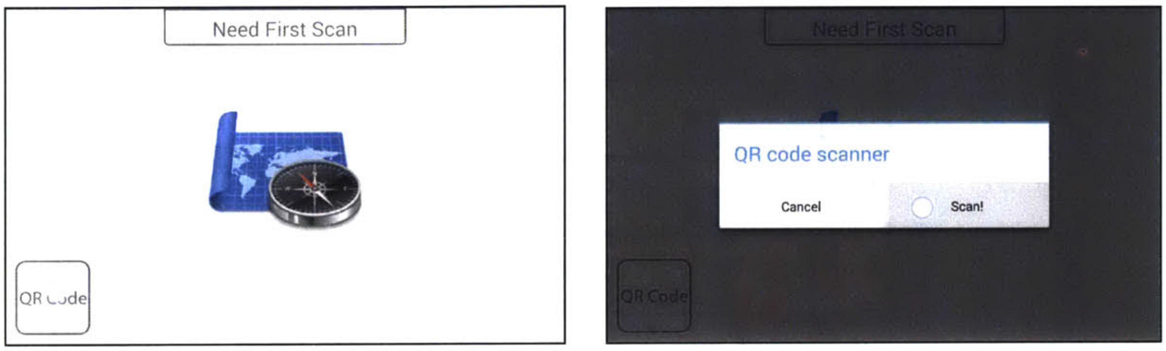


Figure 7: QR Code Scan View

To scan the QR code, you will first select the **QR Code** button on the bottom left by pressing on it. This brings up the Barcode Scanner app and allows you to scan QR codes. Aim the camera at the QR code posted on the wall. After you have successfully scanned the QR code, you will be returned to the Top View.

Running a Navigation Trial

Do not press the power button while using the application

Running a Computer-Selected Navigation Trial

The computer-selected mode will perform all route planning functions to include rerouting due to updates and deviations from your path. You will be informed of all route changes and updates by the app.

To begin the trial, you will scan the start QR code to set that as your start location. See Understanding the QR Code Scan View for information on this process.

Your trial time begins when you press Begin after the first QR Code scan is completed (refer to Figure 5 or 6 to see the view in Computer-selected mode).

The first map you will be able to see after selecting Begin will be the **Top View.** See Understanding the Top View above for more information. In addition see Understanding the Side View above for information about that view.

There is also an **Alert Bar** available at the top of the display. You can toggle the Alert Bar on and off by pressing the tab in the upper left hand corner. The Alert Bar includes information about the trial to include navigation updates. See **Figure 7** for an example of the alert bar.

If a new navigation route is calculated for you, a sound notification will play and a message will appear in the Alert Bar at the top of the display. Your new route might not include the same floor changes and QR codes so it is recommended that you toggle to the Side View to view the new route. There will be three alert noises: update, route modifications, and increased point totals. An alert message will accompany each of these. The experimenter will provide an example of each sound.

Running a User-Choice Navigation Trial

Do not press the power button while using the application

In the user-choice mode you will select the route that you wish to use for the trial. The user-choice mode will not change the navigation route without you manually selecting a new option. If you wander too far off of your selected route, then you may be prompted to continue your current route or update to a new shorter route from your location.

To begin the trial, you will scan the start QR code to set that as your start location. See Understanding the QR Code Scan View above for information on this process.

Your trial time doesn't begin until you choose a first route to use for navigation.

You will be given three different route options to choose from to start the trial, shown in the User-choice Side View. Each route option will pick up a certain number of QR Code points and also have a time estimate associated with it. Remember that you only have 180 seconds to complete the trial, so a longer route will be more difficult to complete within the time limit. In addition remember that you will be deducted half of your points for exceeding 4.5 mph.

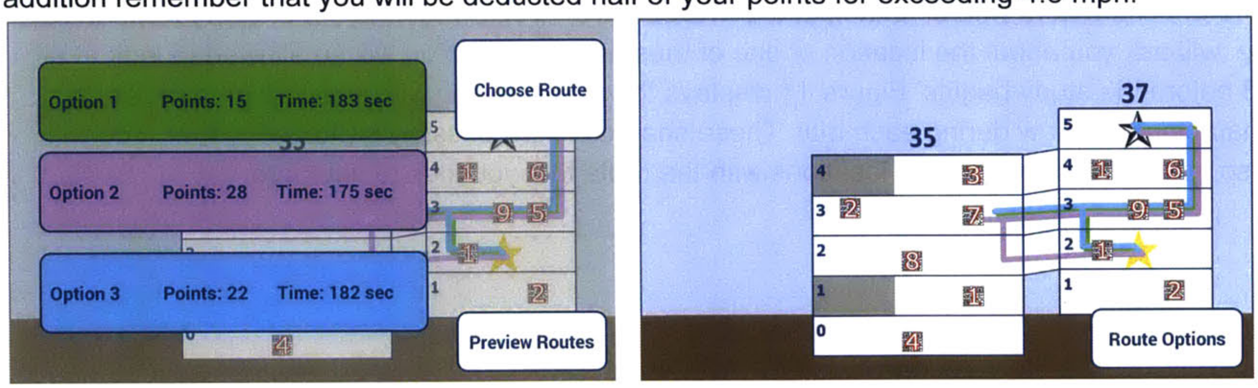


Figure 8: Side View with User-choice route options

The Side View is shown behind the route options and can be viewed by pressing the **Preview Routes** button in the lower right hand of the screen. Each route is color coded with the color of the respective route option buttons. In the above figure for example, route option 1 has a green button and is shown with a green path on the Side View.

Select which route you want to use in the trial by clicking on it and then confirm your choice by clicking on the **Choose Route** button.

Your trial time begins as soon as you press the Choose Route button for the first time.

As with computer-selected mode, after you select Choose Route, the first map you will see will be the Top View. See Understanding the Top View above for more information. In addition see Understanding the Side View above for information about that view.

There is also an **Alert Bar** available at the top of the display. You can toggle the Alert Bar on and off by pressing the tab in the upper left hand corner. The Alert Bar includes information about the trial to include navigation updates. In user-choice mode, when the Alert Bar is on, you can click on the lower section that says **Choose Route** to be brought back to the User-choice Side View that you saw at the beginning of the trial and change to a different route option.

Remember you will have 180 seconds to navigate to the goal location and scan the goal QR code.

Situation Awareness Task

Following each trial, we will ask you about the trial just completed to include: unclaimed points available during the trial, the total of unclaimed points in the trial, the effect of a closure on your route, your initial route, and the start and goal locations. In each hallway will be placed one of 13 different characters with an example shown in Figure 10 below. In one out of every three trials we will ask you about the location of one of those characters. You will be allowed to look at all 13 before the study begins. Figure 11 displays the map that you will be asked to mark which characters you saw during each trial. These characters will possibly switch each trial. You will also mark your start and end locations with the route that you took to get there.

Figure 9: The Situation Awareness Items

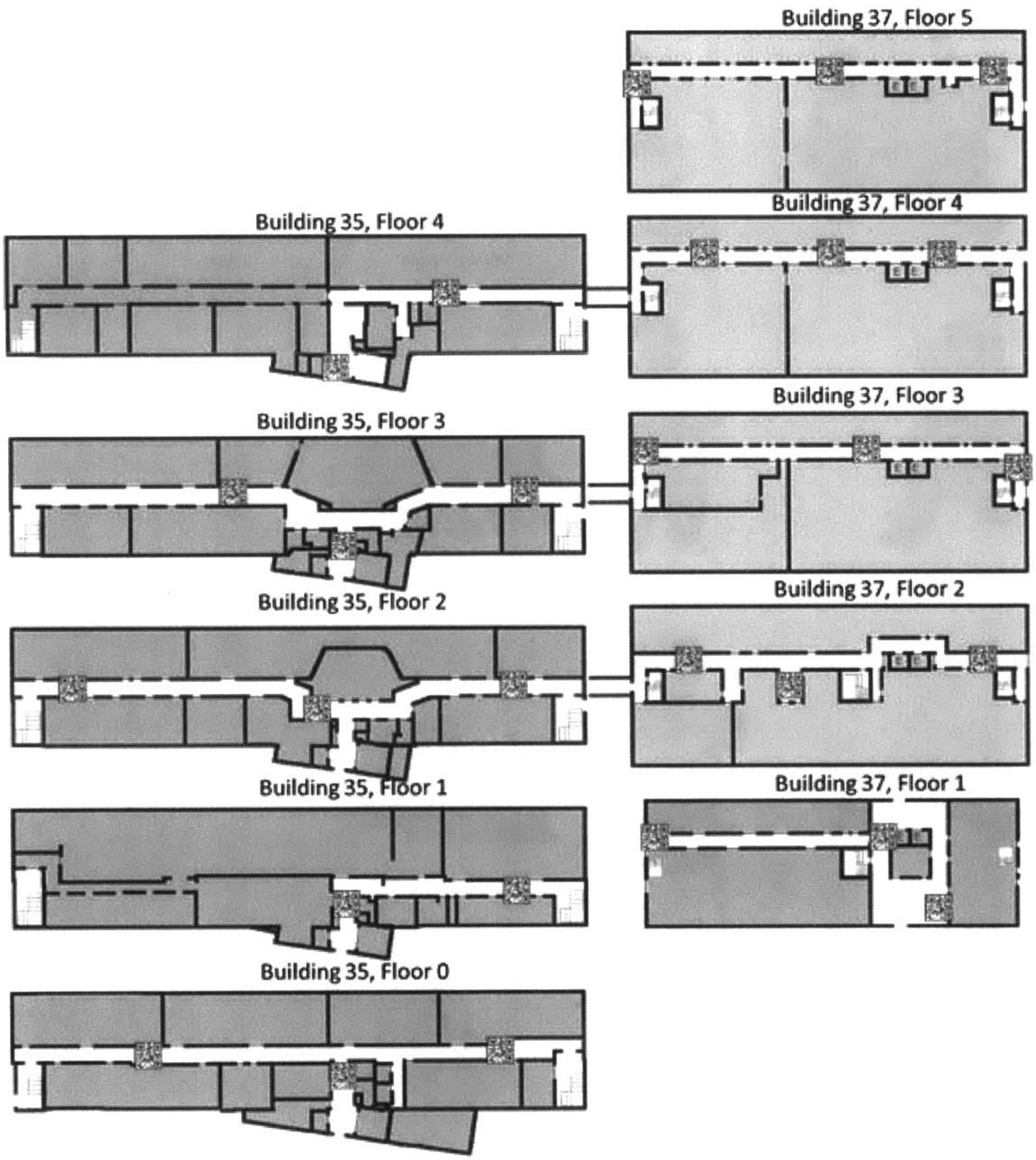


Figure 10: Situation Awareness Question

Readiness Performance Questions

1. What is the difference between the user-choice and computer-selected modes?
2. In user-choice mode, how do you change which route you want to be shown? Explain and then demonstrate.
3. How much time do you have in a trial to reach and scan the goal QR code? When does this time begin?
4. How do you collect points from QR codes posted on the wall? Explain and then demonstrate.
5. In order to earn the most points, should you attempt to scan every QR code that you pass during your trial?
6. In which view, side or top, are you able to see how much time is remaining in the trial? Name a couple other differences between the two views.

Readiness Performance Metrics

1. Understands the difference between user-choice and computer-selected trials when it comes to having the ability to choose the route to follow amongst options or not:
   a. What is the difference between the user-choice and computer-selected modes?

      Computer-selected: The navigation system provides an updated route without user input
      User-choice: User must choose one of three routes to use or update with

2. Knows how to change between route options while completing an user-choice trial:
   a. In user-choice mode, how do you change which route you want to be shown? Explain and then demonstrate.

      Routes are updated in user-choice mode by opening up the alert bar and pressing the lower button across the top of the screen that says "Route Options"
      The alert bar can be opened by the toggle button in the upper left corner

3. Knows when the trial time is started in each trial and how much time they have to make it to the goal QR code:
   a. How much time do you have in a trial to reach and scan the goal QR code? When does this time begin?

      180 seconds in each trial to scan the goal QR code
      Computer-selected trial time starts after pressing the Begin button following your first QR scan
      User-choice trial time starts after your first route choice following your first QR scan

4. Knows how to scan a QR code posted on the wall:
   a. How do you collect points from QR codes posted on the wall? Explain and then demonstrate.

      The Scan QR button is on the lower left when in Top View mode
      The Top View can be accessed from the Side View by first pressing the button in the lower right that says "Top View"

5. Understands the time restrictions and penalties associated with collecting the maximum number of points :
   a. In order to earn the most points, should you attempt to scan every QR code that you pass during your trial?

Not necessarily, scanning QR codes takes time and you want to maximize your points in the 3 minute time window. In addition, there are penalties for being late that will significantly outweigh one QR code.

6. Understands the difference between Side View and Top View :
   a. In which view, side or top, are you able to see how much time is remaining in the trial? Name a couple other differences between the two views.

      The Top View displays how much time is remaining in the trial. The Top View shows the local floor, while the Side View shows the global map. The Top View also displays: points collected, remaining QR codes on your route and remaining points on your route. You can only access QR Code button from Top View.

Name	Trial	Update	Start	Goal
Alpha	1	none	37:2:106	35:4:113
Bravo	1	close 37:5 (60 sec)	37:2:106	35:4:113
Charlie	2	none	37:5:59	35:2:132
Delta	2	close 35:3 (30 sec)	37:5:59	35:2:132
Echo	3	none	35:0:102	37:4:54
Foxtrot	3	close 35:2 (20 sec)	35:0:102	37:4:54
Golf	4	none	37:3:61	35:0:133
Hotel	4	close 35:2 (60 sec)	37:3:61	35:0:133
India	5	none	35:3:105	37:2:106
Juliett	5	close 37:4 (40 sec)	35:3:105	37:2:106
Kilo	6	none	35:1:109	37:4:61
Lima	6	close 37:2 (35 sec)	35:1:109	37:4:61
Mike	7	none	37:5:59	37:2:106
November	7	close 37:3 (25 sec)	37:5:59	37:2:106
Oscar	8	none	37:1:63	35:2:116
Papa	8	close 37:3 (30 sec)	37:1:63	35:2:116
Quebec (walk)	all 1s	none	anywhere	none

Appendix N

Situation Awareness Characters and Locations

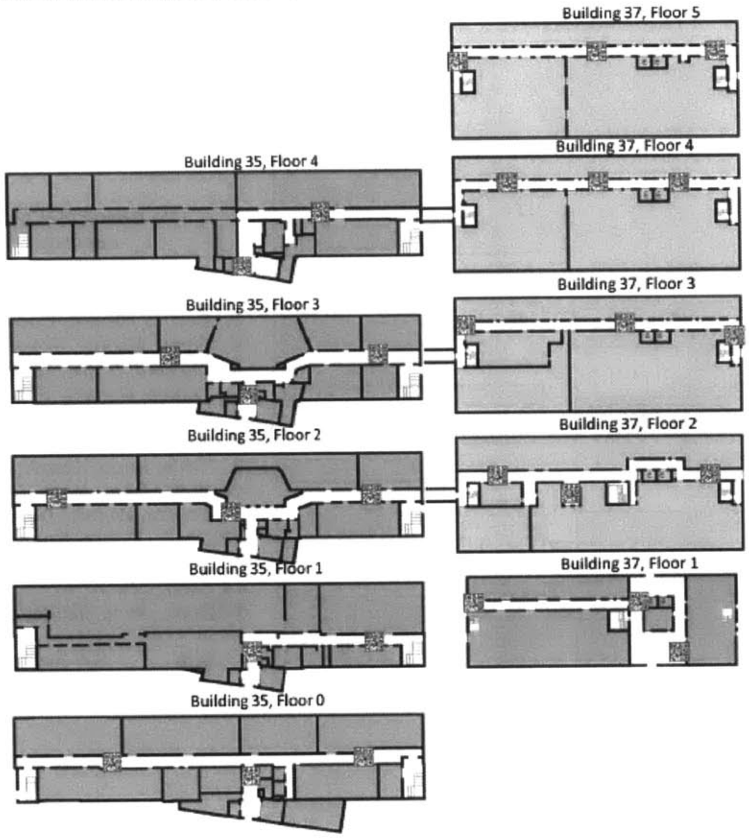
Below are the files that define the SA task throughout the experiment. Figure N-1, shows the 13 characters rotated throughout the hallways and also represents the reference sheet used by subjects when completing the survey shown in Figure N-2. The characters were rotated every three trials and this rotation is shown in Figure N-3. Finally, Figure N-4 shows an example of the scoring of the SA survey to the calculation of the SA composite score.

Figure N-1: Characters used in Secondary SA Test

Subject # __________ Trial #___________ Experimental Day:_________

First Trial:

Mark on this map where the start and goal location were located with a 'S' for start and 'G' for goal, and the floors that you visited with an 'X':



Did you see any characters in the hall? Y / N (circle choice)

If so, where were they? Mark on the map the location with the initials of character as shown in the handout.

How much did you sleep last night? (in hours) __________

Figure N-2: SA Survey

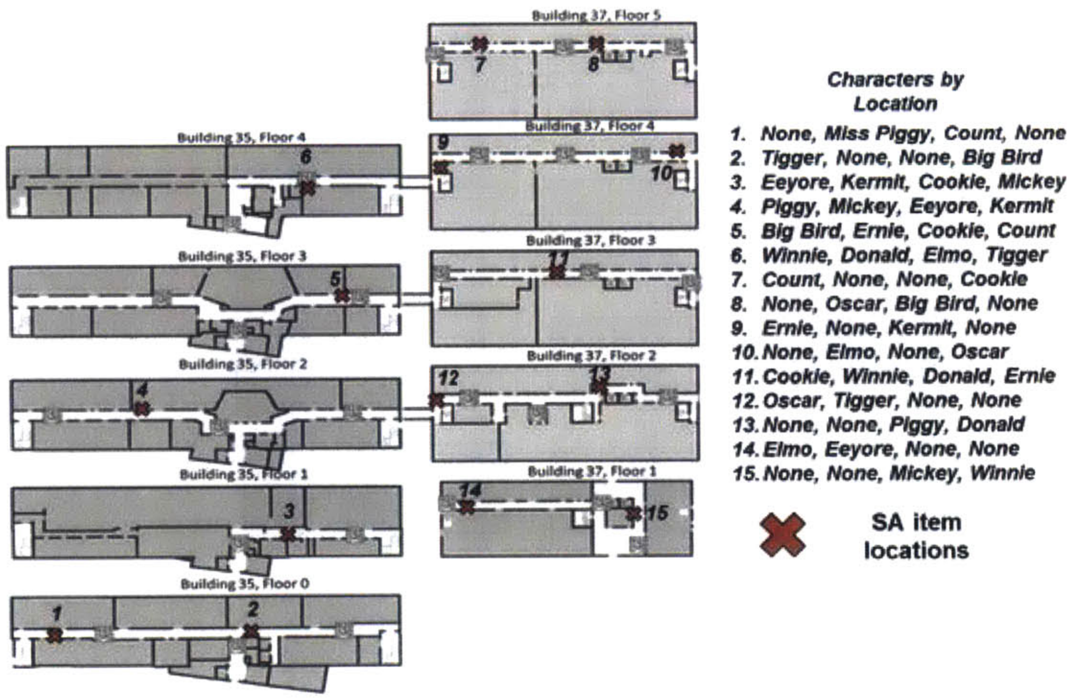
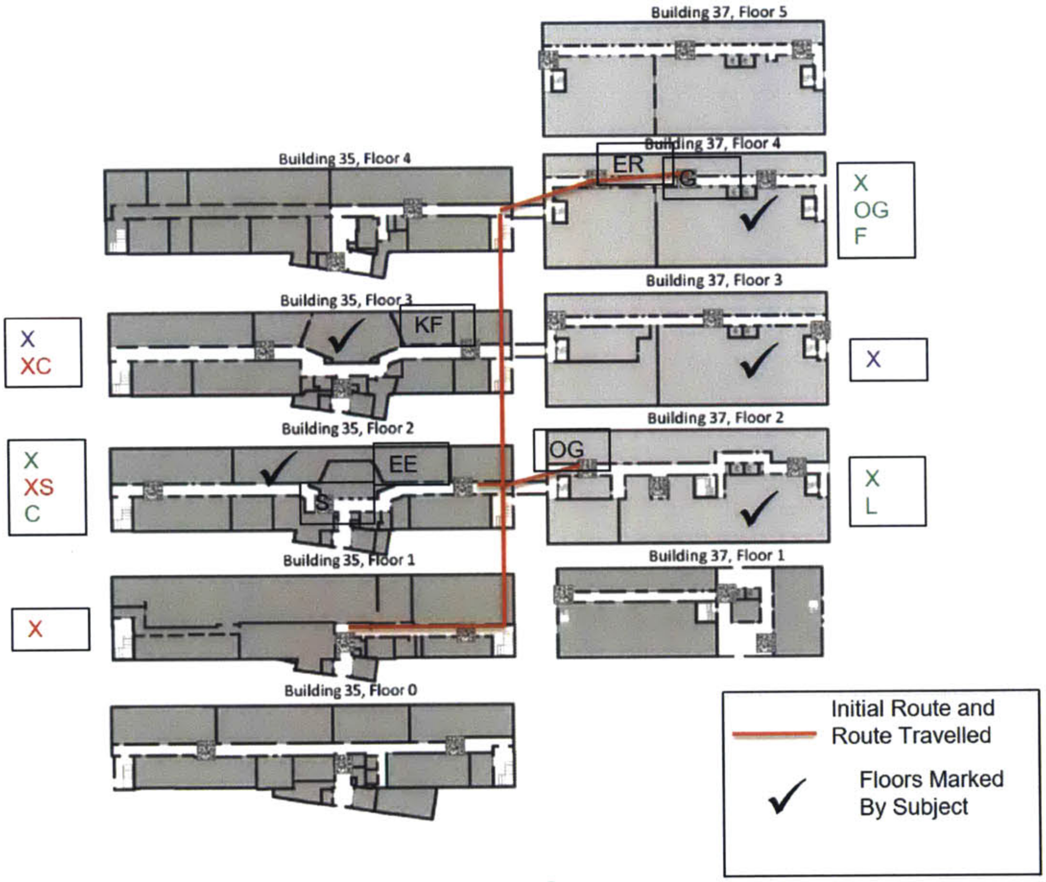


Figure N-3: Location of Characters depending on Trial


Subject # Even Example Trial # 1 (Kilo) Experimental Day: 2

First Trial:

Mark on this map where the start and goal location were located with a 'S' for start and 'G' for goal, and the floors that you visited with an 'X':

Did you see any characters in the hall? (Y) / N (circle choice)

If so, where were they? Mark on the map the location with the initials of character as shown in the handout.

How much did you sleep last night? (in hours)

	Incorrect	Correct	N/A	Score
Start Correct	XS	OS		0
Goal Correct	XG	OG		1
Characters Seen (Y/N)	(N)	(Y)		1
Characters Seen Total	Total of Seen	Total S, F & L		4
Characters Correct Seen	XC	C		3
Characters Correct Floor	Check Seen	F		2
Characters Correct Location	Check Floor	L		1
Floors With Scanned QR Codes	X	X	X	3/4

Composite SA = 14

Figure N-4: Example SA Scoring

Appendix O

Experimental Conditions

Table O.1 shows the two experimental orders applied to subjects.

Table O.1: Treatments

Odd Subjects					Even Subjects				
Block	Block Mode	Trial	Update	Trial Call Sign	Block	Block Mode	Trial	Update	Trial Call Sign
1	Computer-select	5	Y	Lima	2	User-choice	6	Y	Papa
1	Computer-select	3	N	Golf	2	User-choice	5	N	Kilo
1	Computer-select	1	Y	Bravo	2	User-choice	4	Y	Juliett
1	Computer-select	6	Y	Papa	2	User-choice	3	N	Golf
1	Computer-select	4	N	India	2	User-choice	2	Y	Delta
1	Computer-select	2	N	Charlie	2	User-choice	1	N	Alpha
2	User-choice	1	N	Alpha	1	Computer-select	2	N	Charlie
2	User-choice	2	Y	Delta	1	Computer-select	4	N	India
2	User-choice	3	N	Golf	1	Computer-select	6	Y	Papa
2	User-choice	4	Y	Juliett	1	Computer-select	1	Y	Bravo
2	User-choice	5	N	Kilo	1	Computer-select	3	N	Golf
2	User-choice	6	Y	Papa	1	Computer-select	5	Y	Lima
4	User-choice	6	N	Oscar	3	Computer-select	5	N	Kilo
4	User-choice	5	Y	Lima	3	Computer-select	3	Y	Hotel
4	User-choice	4	N	India	3	Computer-select	1	N	Alpha
4	User-choice	3	Y	Hotel	3	Computer-select	6	N	Oscar
4	User-choice	2	N	Charlie	3	Computer-select	4	Y	Juliett
4	User-choice	1	Y	Bravo	3	Computer-select	2	Y	Delta
3	Computer-select	2	Y	Delta	4	User-choice	1	Y	Bravo
3	Computer-select	4	Y	Juliett	4	User-choice	2	N	Charlie
3	Computer-select	6	N	Oscar	4	User-choice	3	Y	Hotel
3	Computer-select	1	N	Alpha	4	User-choice	4	N	India
3	Computer-select	3	Y	Hotel	4	User-choice	5	Y	Lima
3	Computer-select	5	N	Kilo	4	User-choice	6	N	Oscar

Appendix P

Linear Regression Figures

Below are the figures showing the significance for investigated main and interaction effects from regression only. All error bars signify standard deviation. Grey lines represent subject means and standard deviations while black lines are the group means and standard deviations.

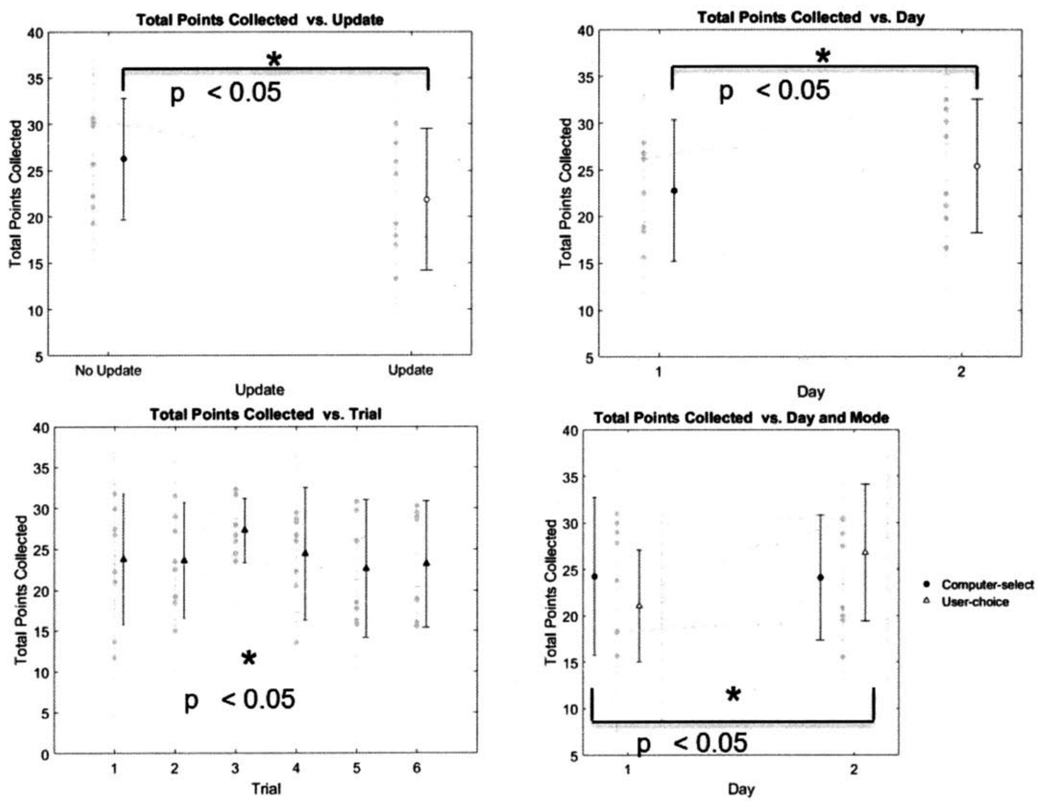


Figure P-1: Total Points Collected Before Penalties refer to Table 4.3

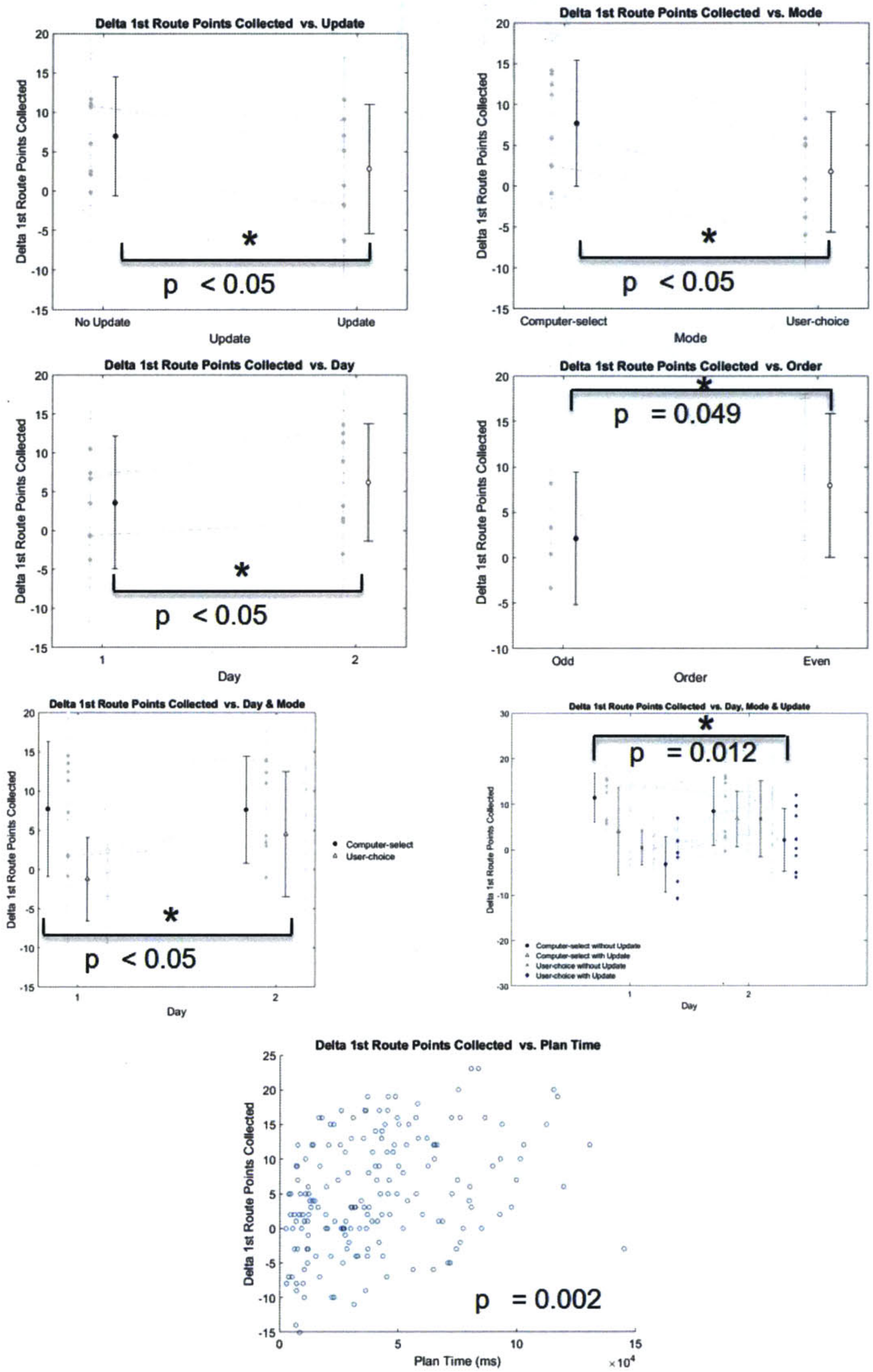


Figure P-2: Delta 1st Route Points Before Penalties refer to Table 4.4

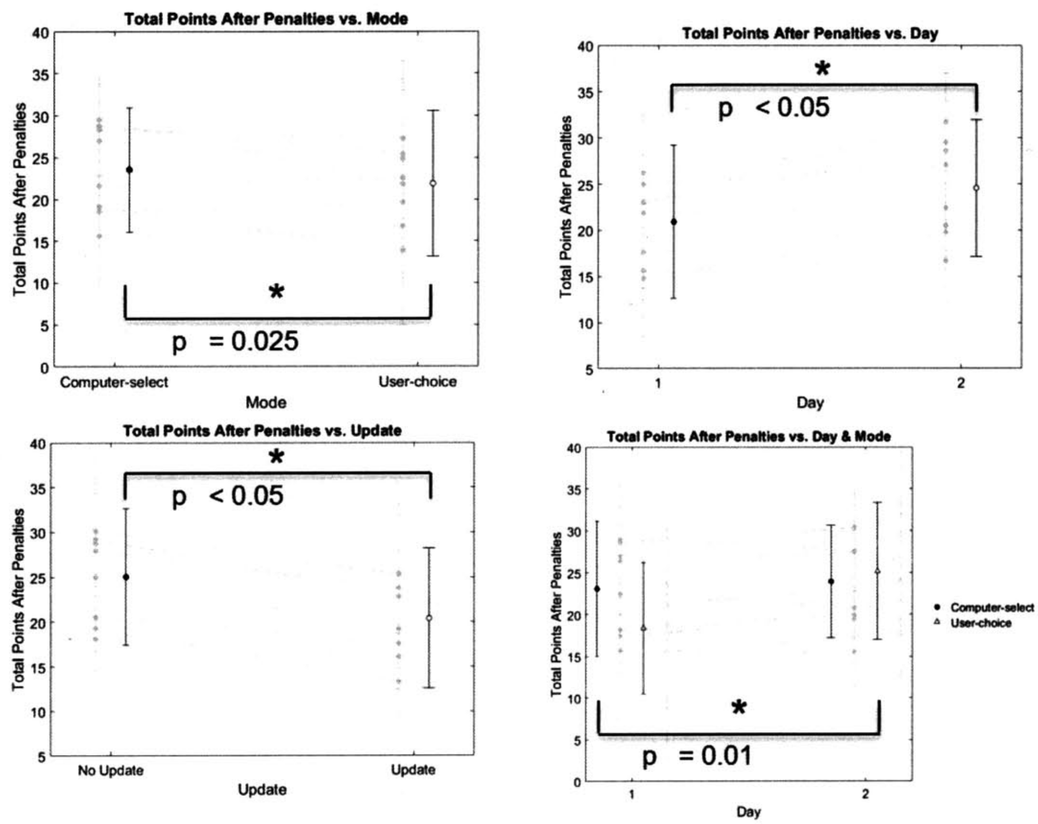


Figure P-3: Total Points After Penalties refer to Table 4.6

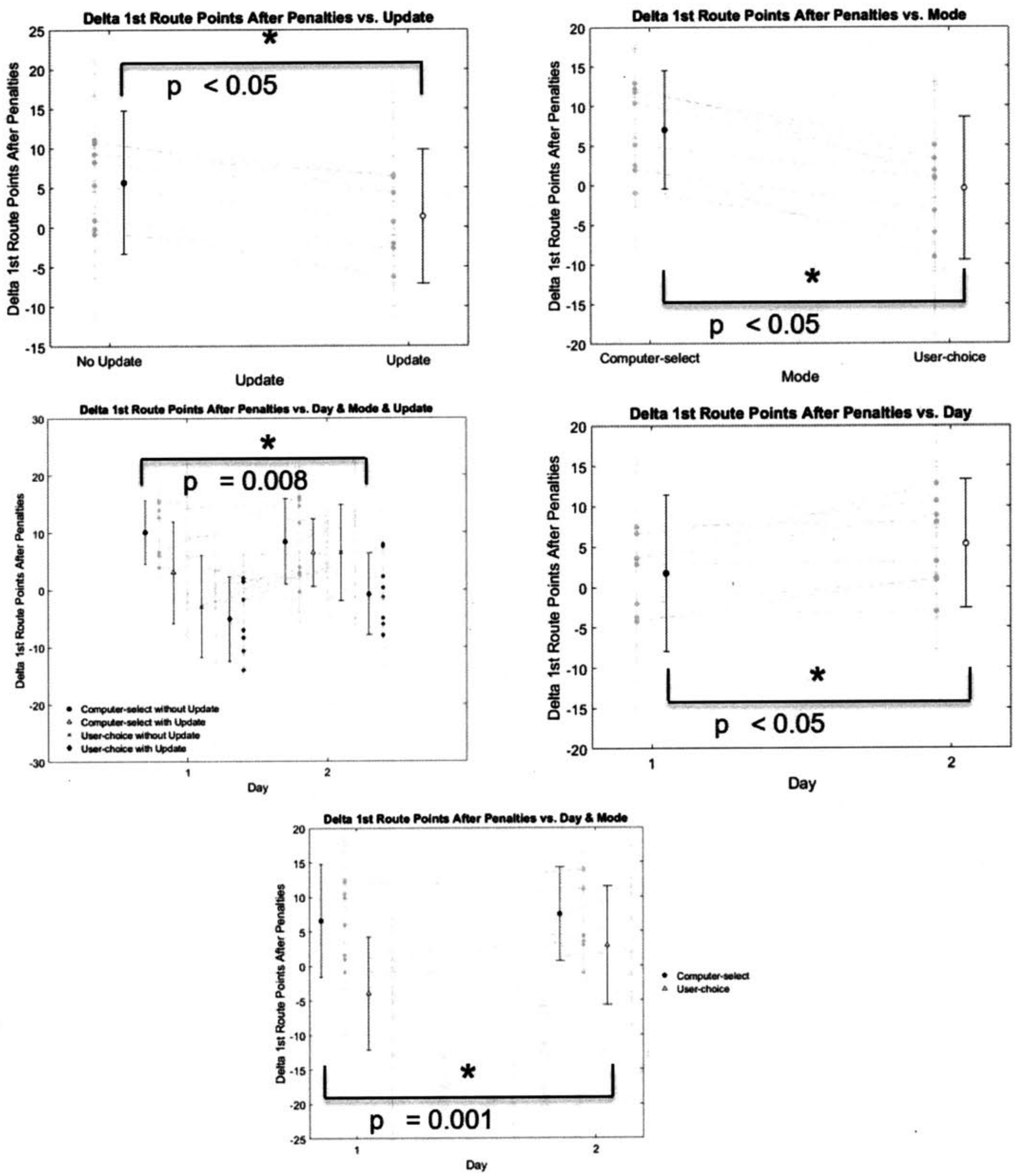


Figure P-4: Delta 1st Route Points After Penalties refer to Table 4.7

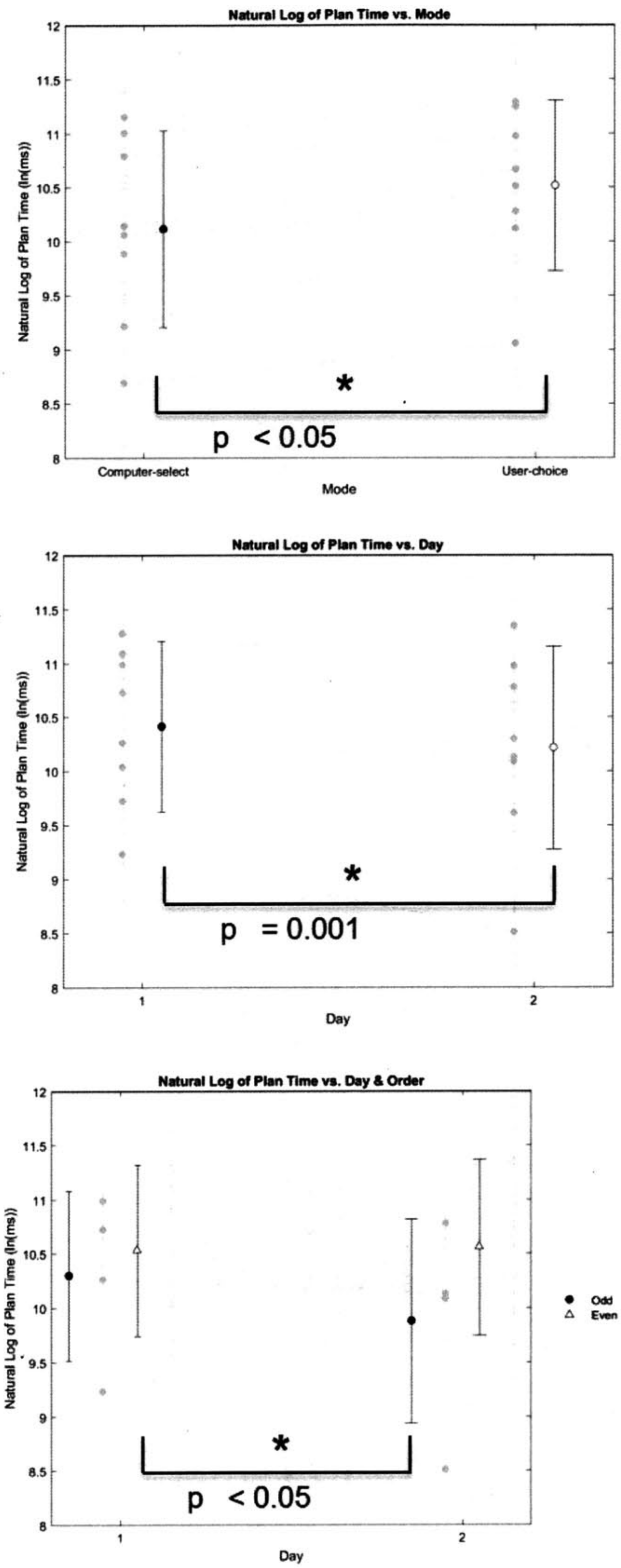


Figure P-5: Natural Log of Planning Time refer to Table 4.8

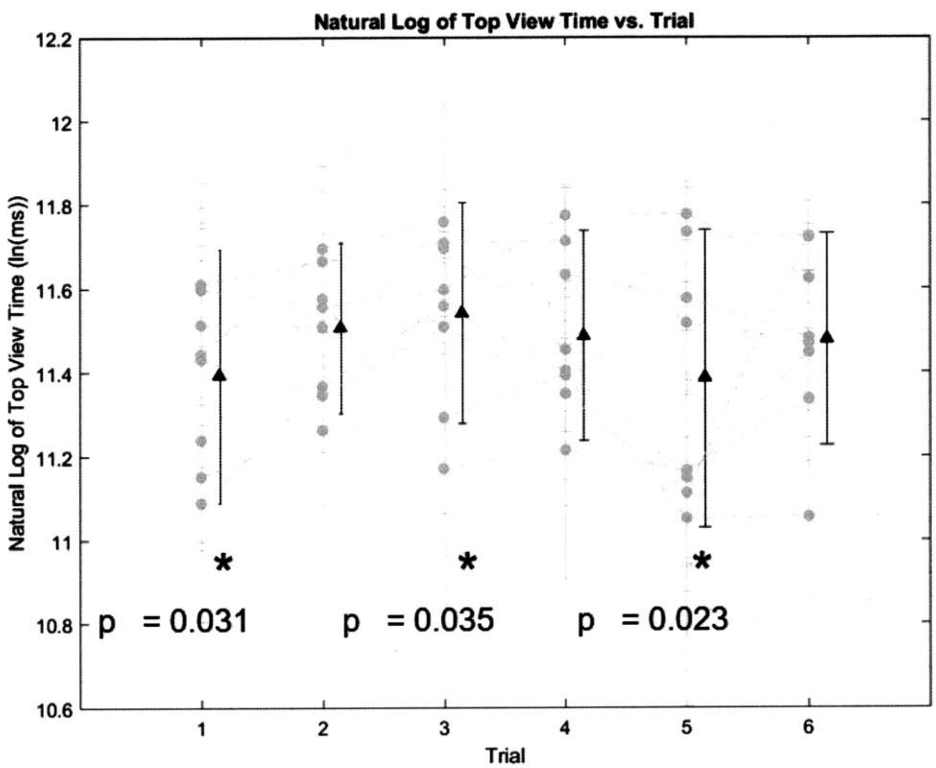
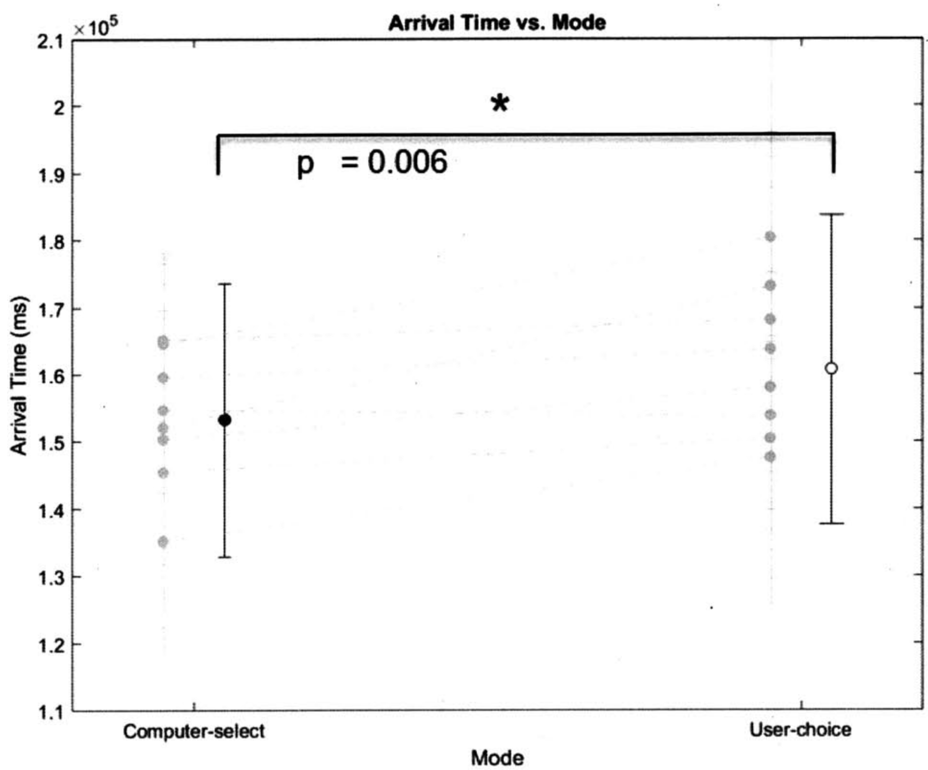


Figure P-6: Natural Log of Top View Time refer to Table 4.9



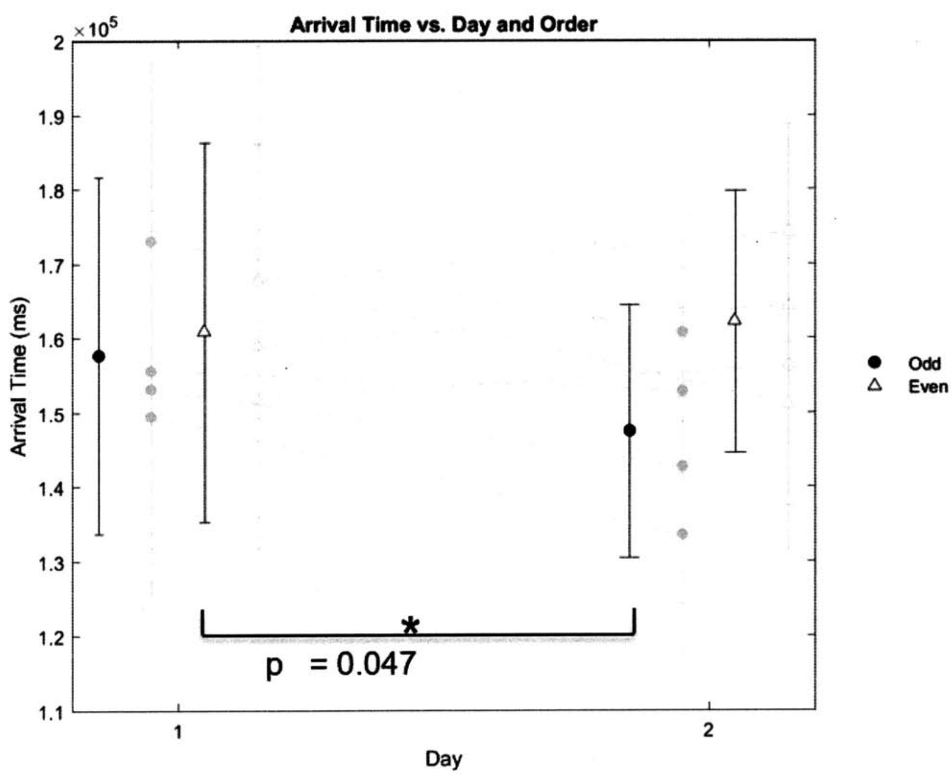


Figure P-7: Arrival Time refer to Table 4.10

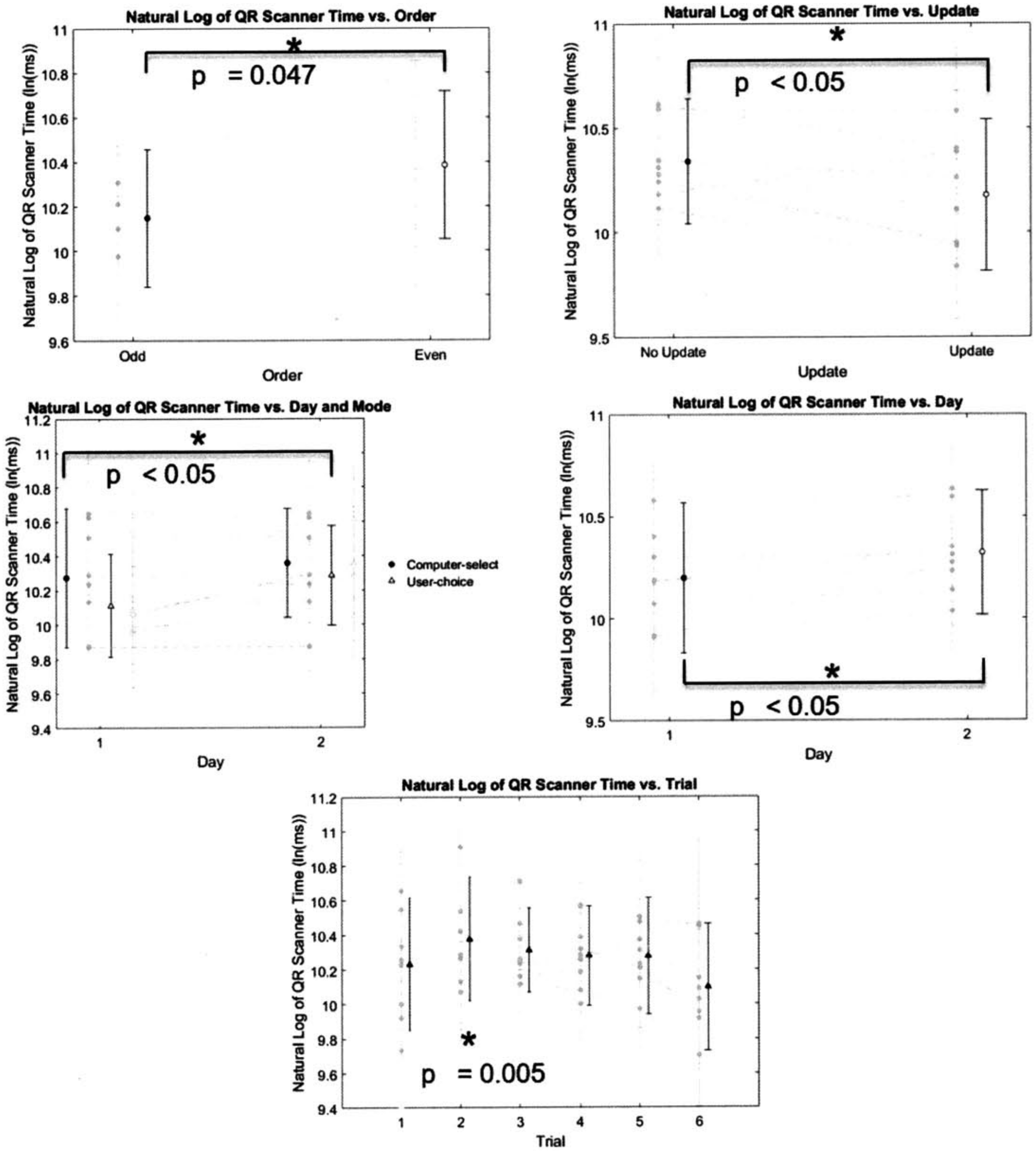


Figure P-8: Natural Log of QR Code Scanner Time refer to Table 4.11

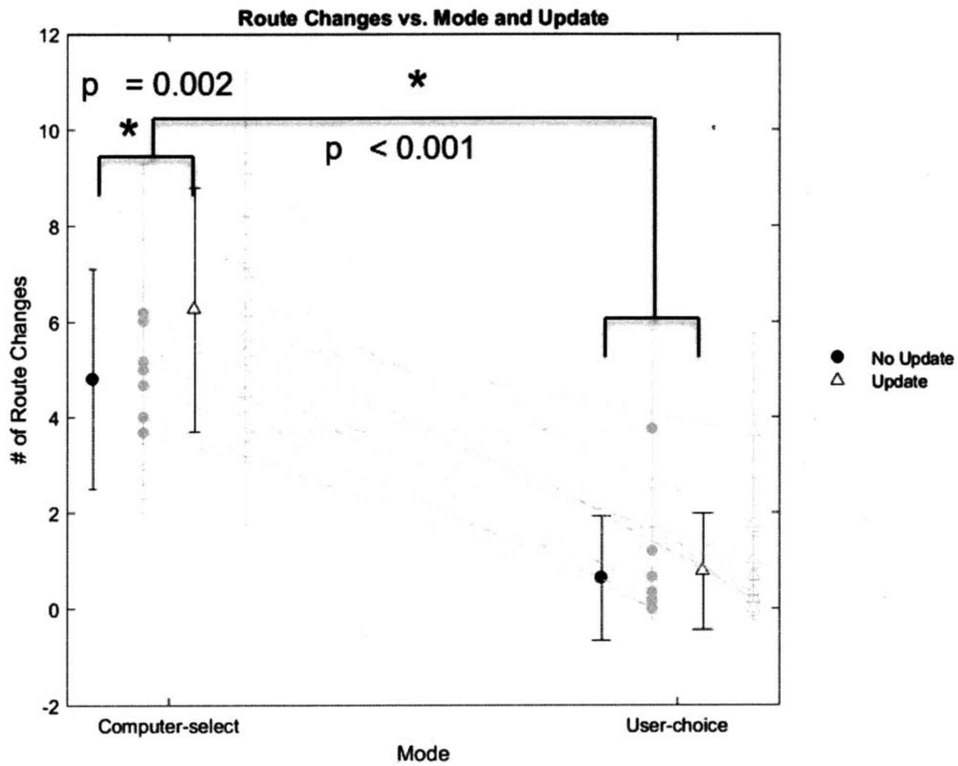


Figure P-9: Route Changes refer to Table 4.12

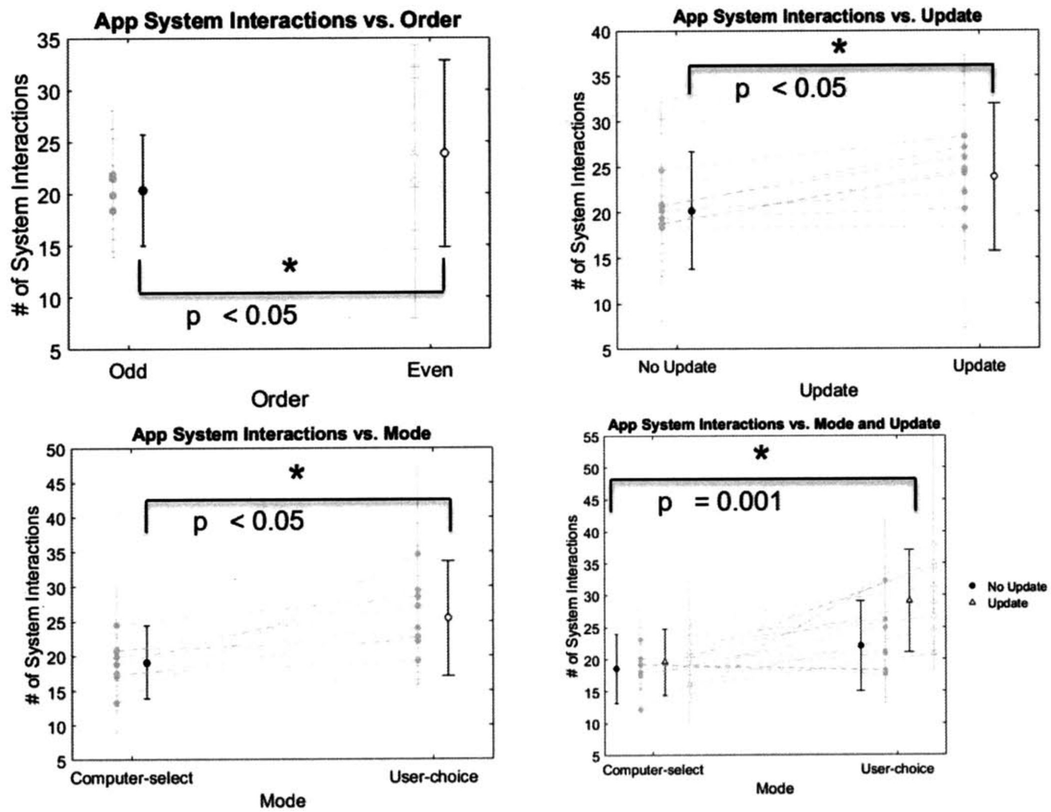


Figure P-10: System Interactions refer to Table 4.13

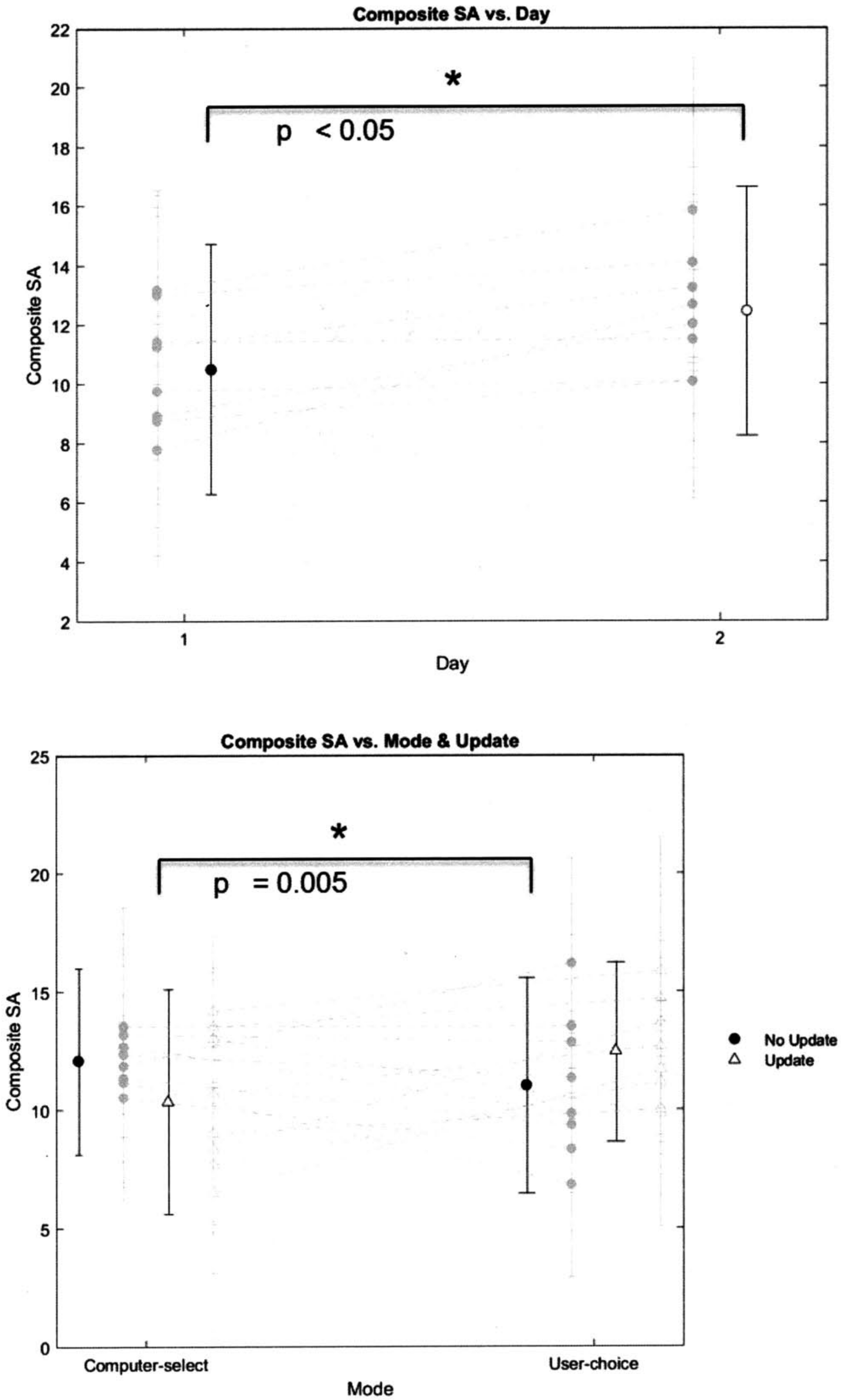


Figure P-11: Composite SA Linear refer to Table 4.15

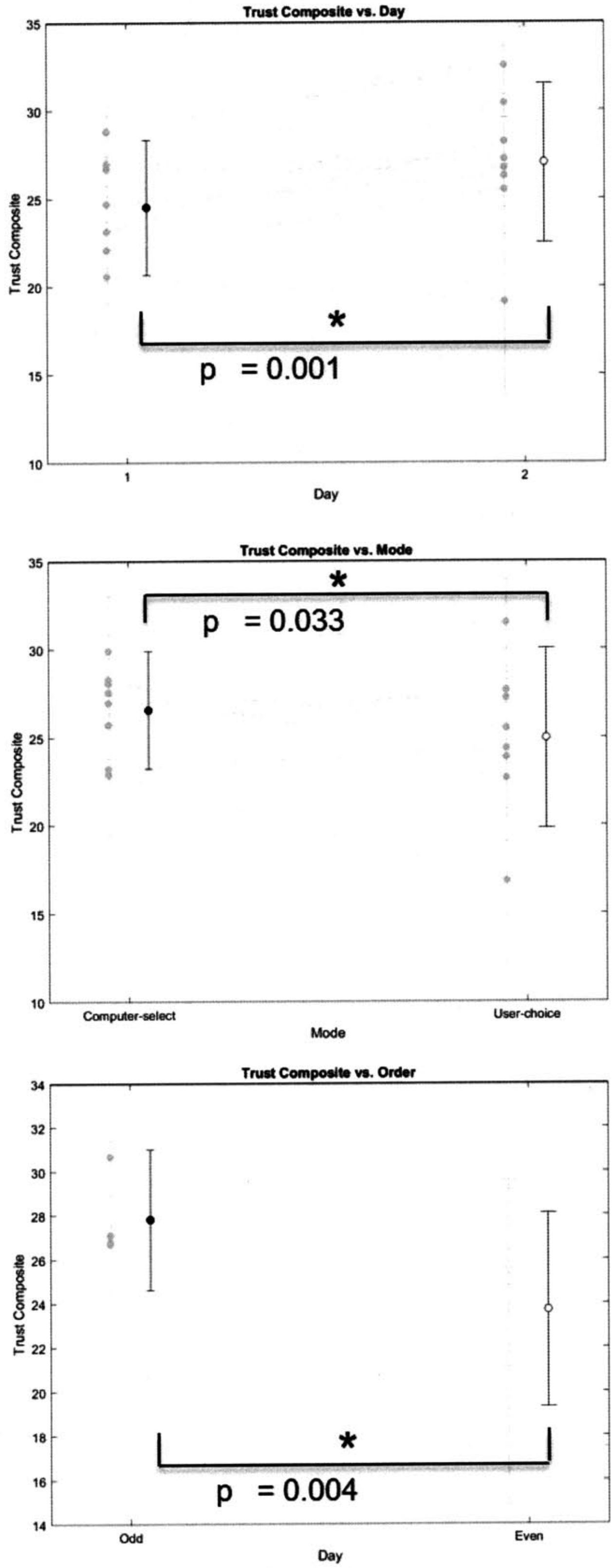


Figure P-12: Composite Trust refer to Table 4.18

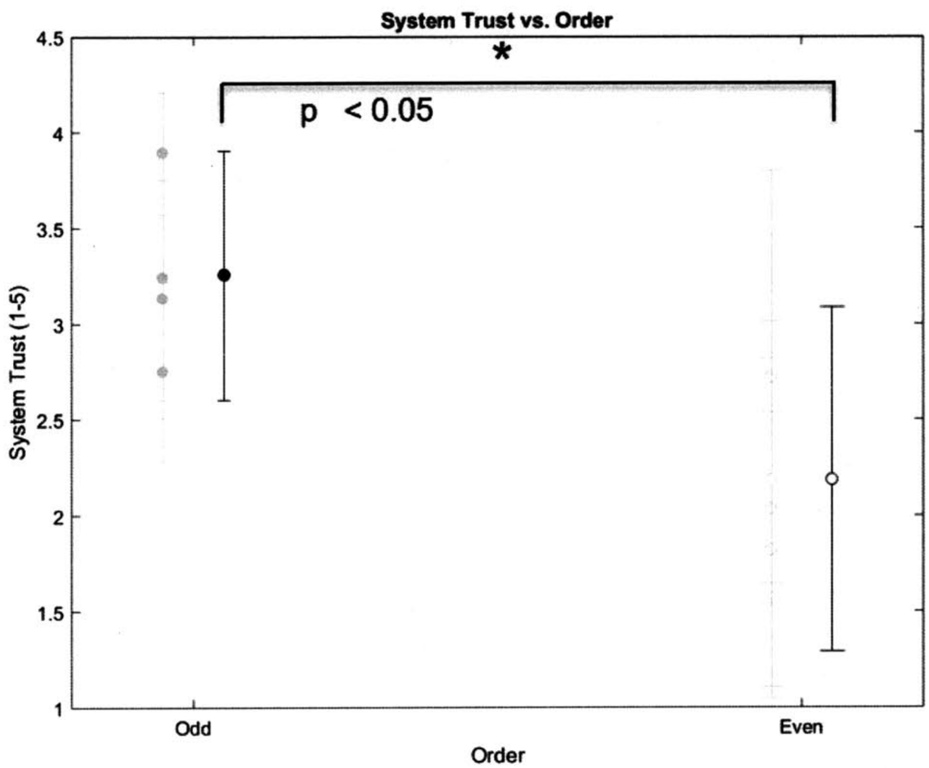


Figure P-13: System Trust refer to Table 4.19

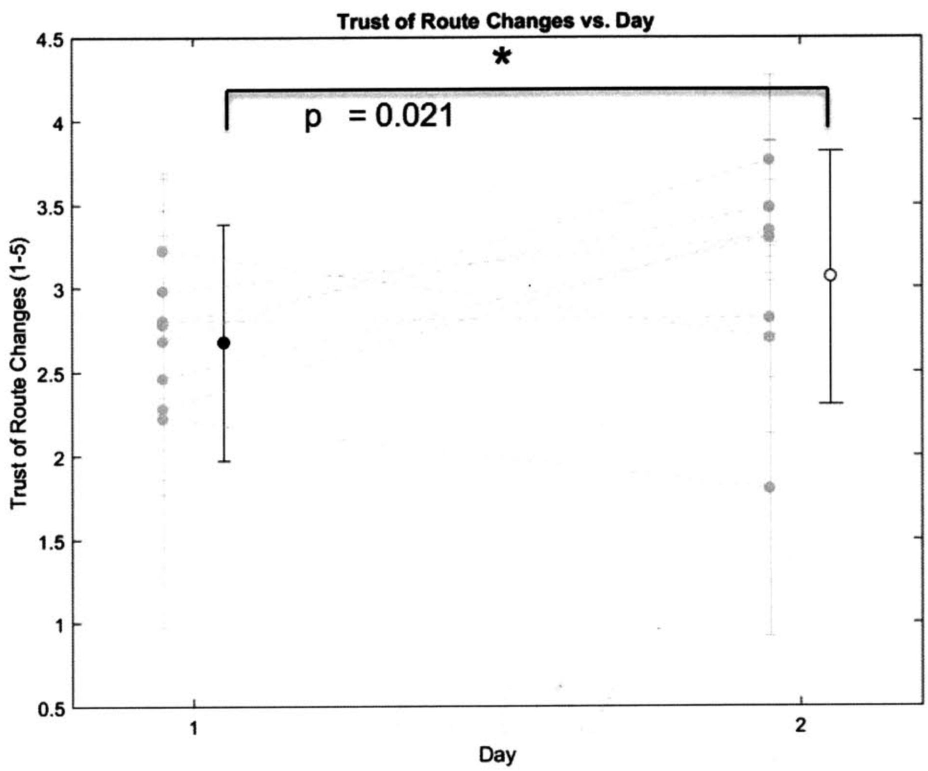


Figure P-14: Trust of Route Changes refer to Table 4.20

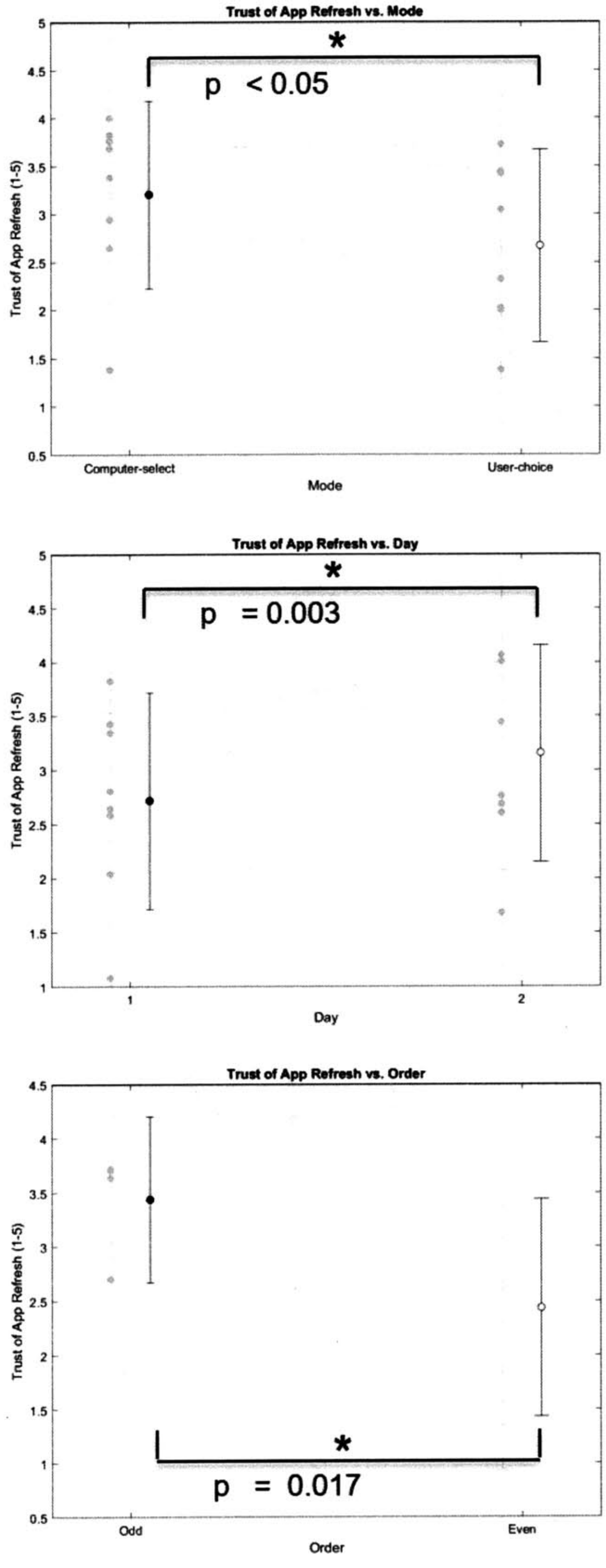


Figure P-15: App Refresh Reliability refer to Table 4.21

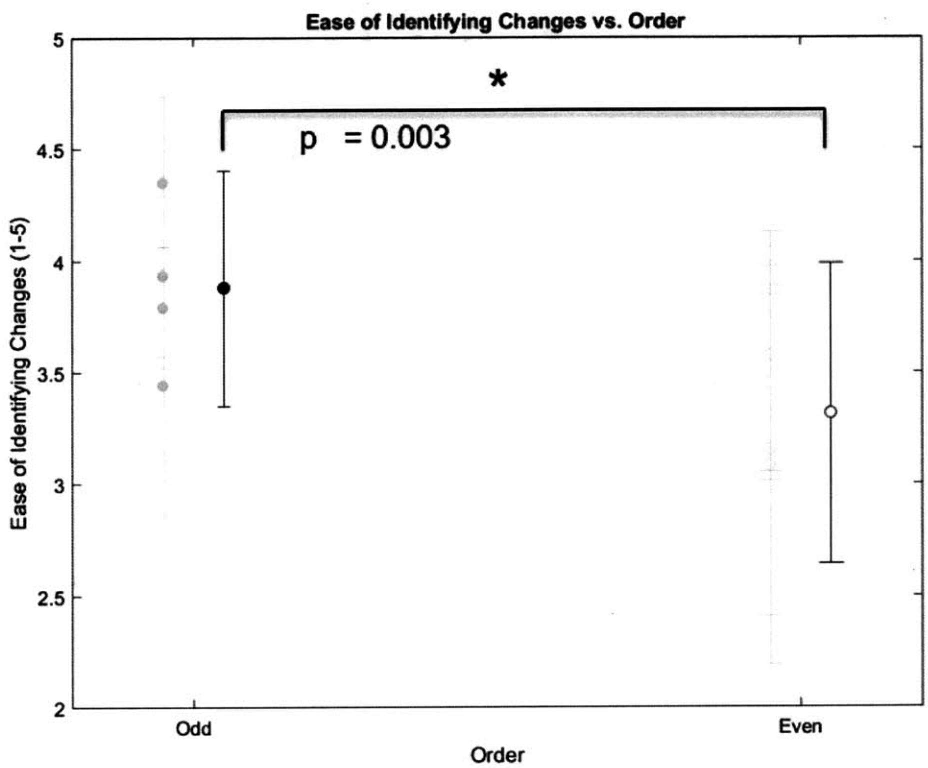
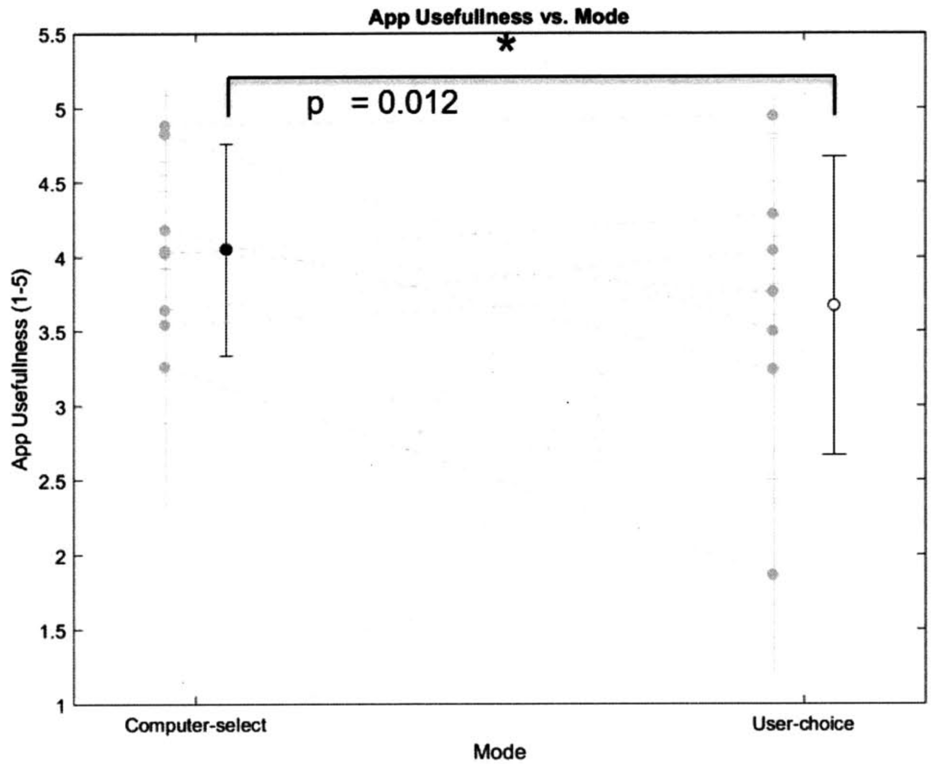


Figure P-16: Ease of Identifying Route Changes refer to Table 4.22



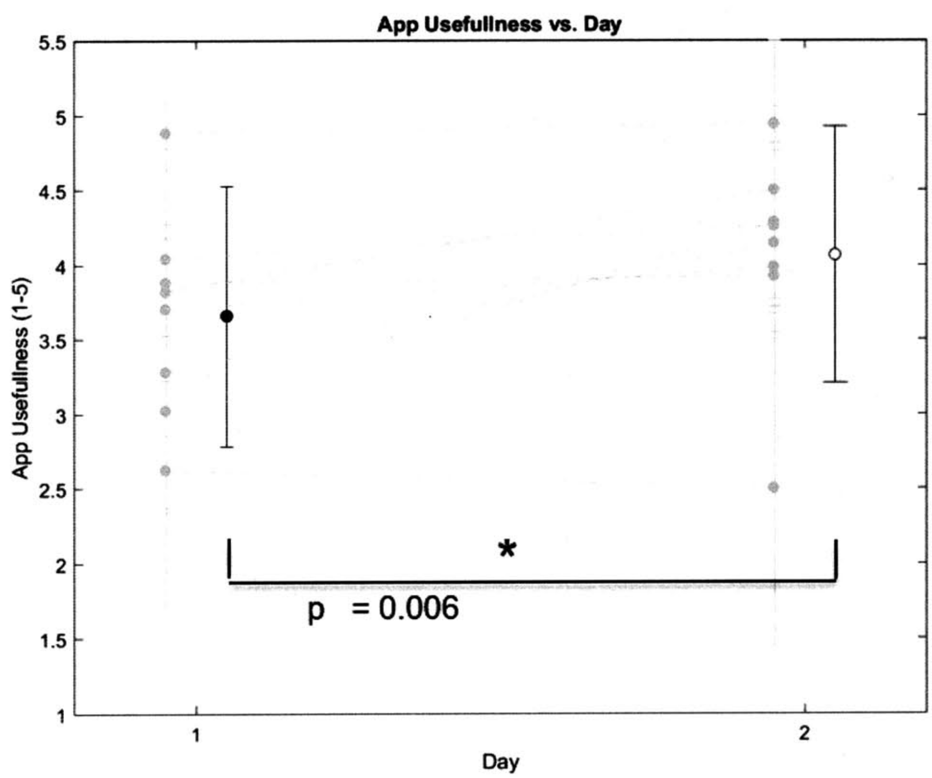


Figure P-17: Usefulness of the App refer to Table 4.23

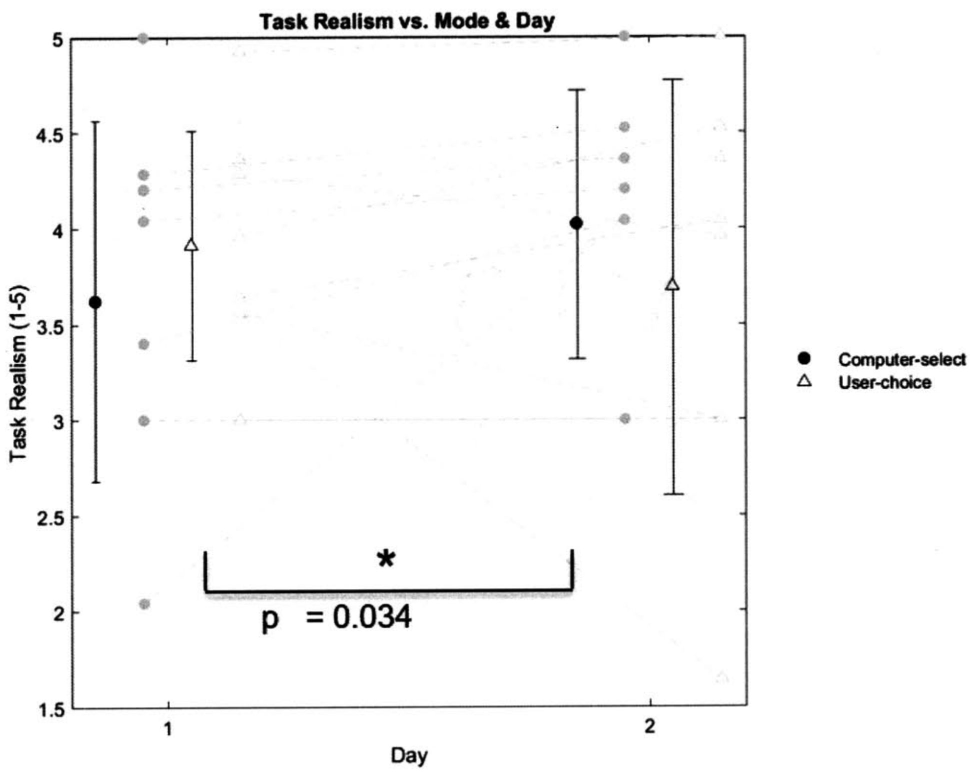


Figure P-18: Task Realism refer to Table 4.25

Appendix Q

Explanation of Mixed Regressions

This appendix shows examples of how to interpret the mixed regression models for the overall scores. The regression table is shown as well as a graphical interpretation of the impact of each model item.

Q.1 Total Points Collected Before Penalties

Table Q.1 can be seen graphically in Figures Q-1 to Q-6 with the itemized impact of each model element by trial.

Table Q.1: Total Points Collected Before Penalties

Variable	$\hat{\beta}$	SE	Z	p
Intercept	24.04	1.75	13.74	<0.001
Day	-1.42	0.34	-4.15	<0.001
Update	2.12	0.34	6.23	<0.001
Trial 1	-0.69	0.76	-0.91	0.363
Trial 2	-0.73	0.76	-0.96	0.336
Trial 3	3.76	0.77	4.86	<0.001
Trial 4	-0.01	0.78	-0.12	0.990
Trial 5	-1.45	0.74	-1.96	0.051
Mode*Day	1.47	0.34	4.30	<0.001

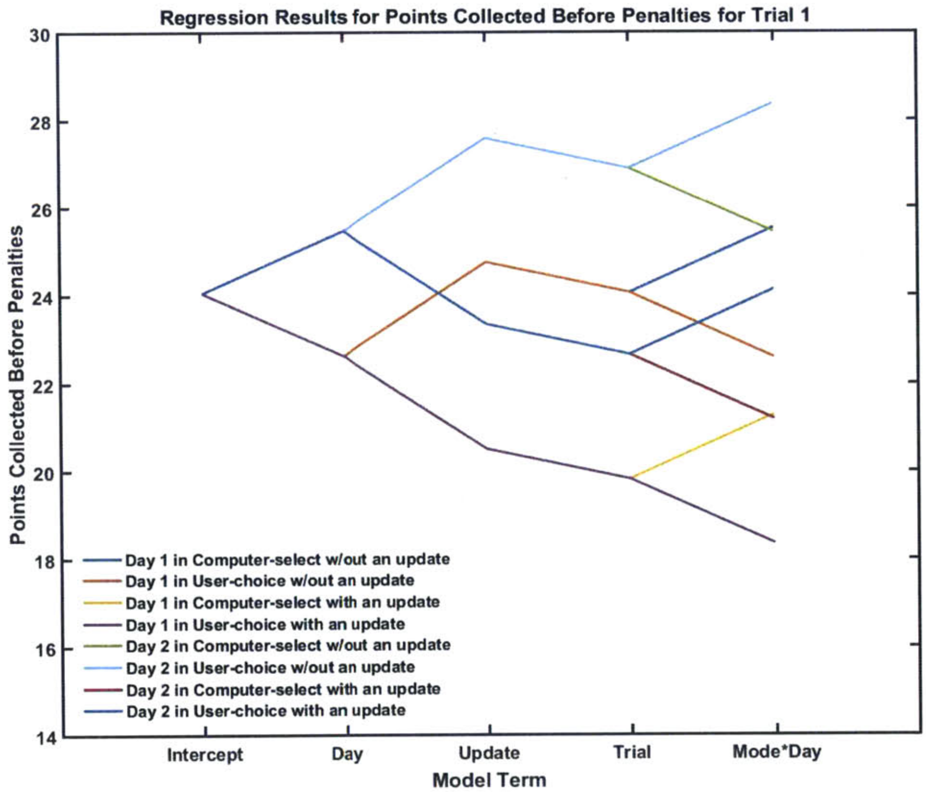


Figure Q-1: Impacts of the Individual Elements on the Linear Regression Model for Points Collected before Penalties for Trial 1

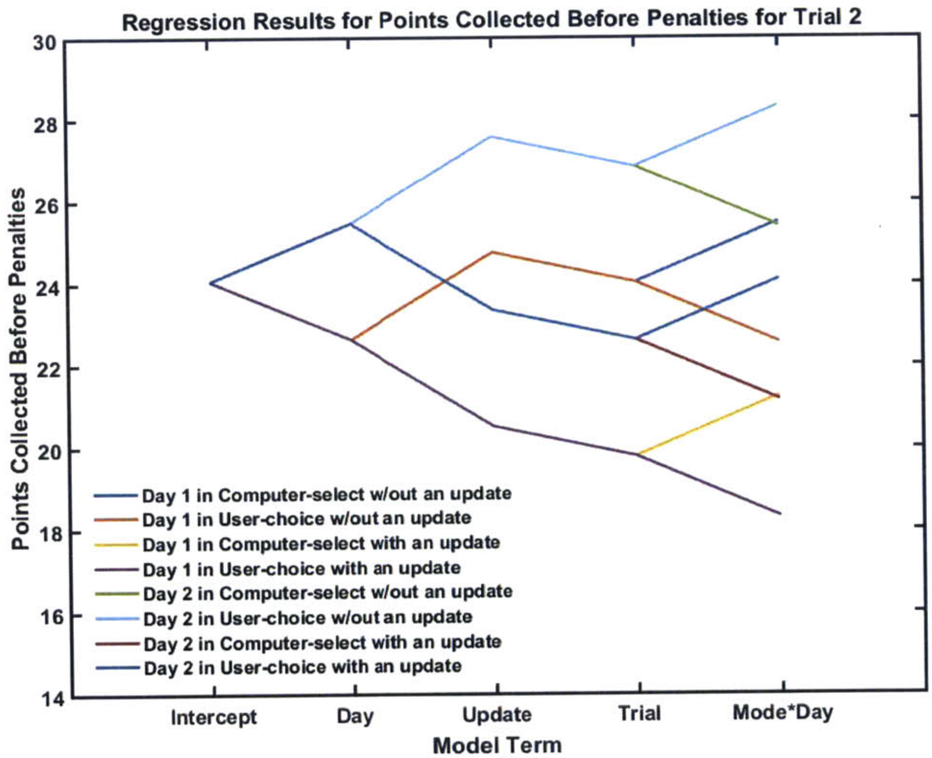


Figure Q-2: Impacts of the Individual Elements on the Linear Regression Model for Points Collected before Penalties for Trial 2

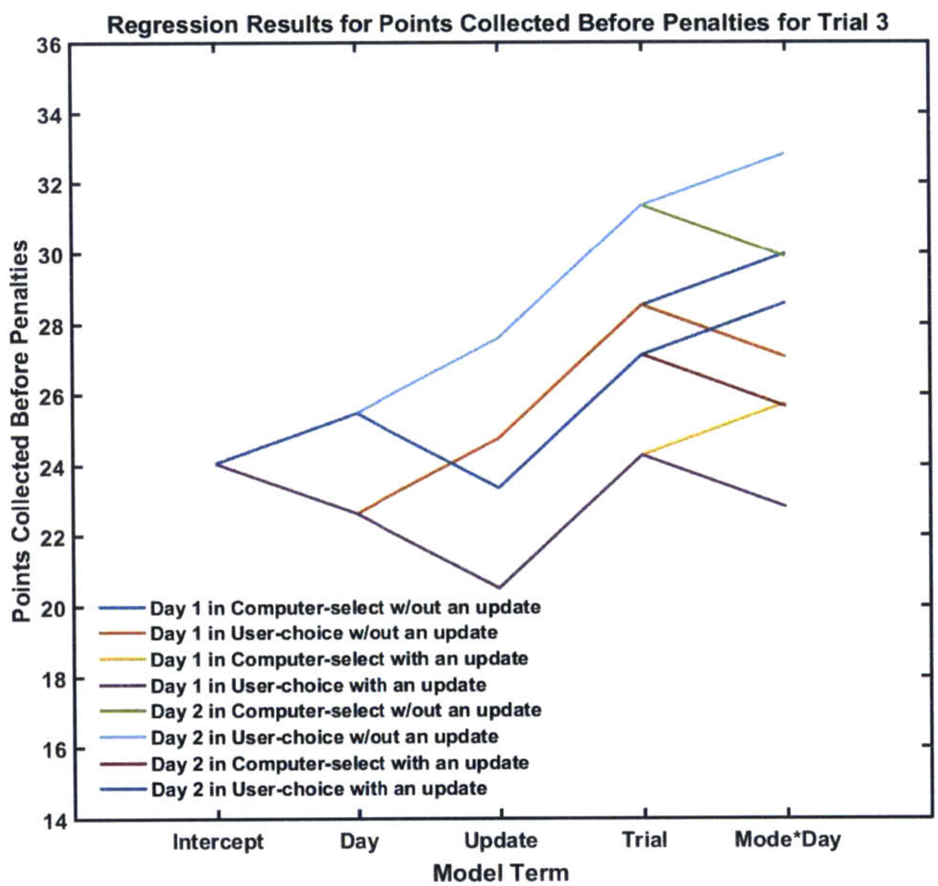


Figure Q-3: Impacts of the Individual Elements on the Linear Regression Model for Points Collected before Penalties for Trial 3

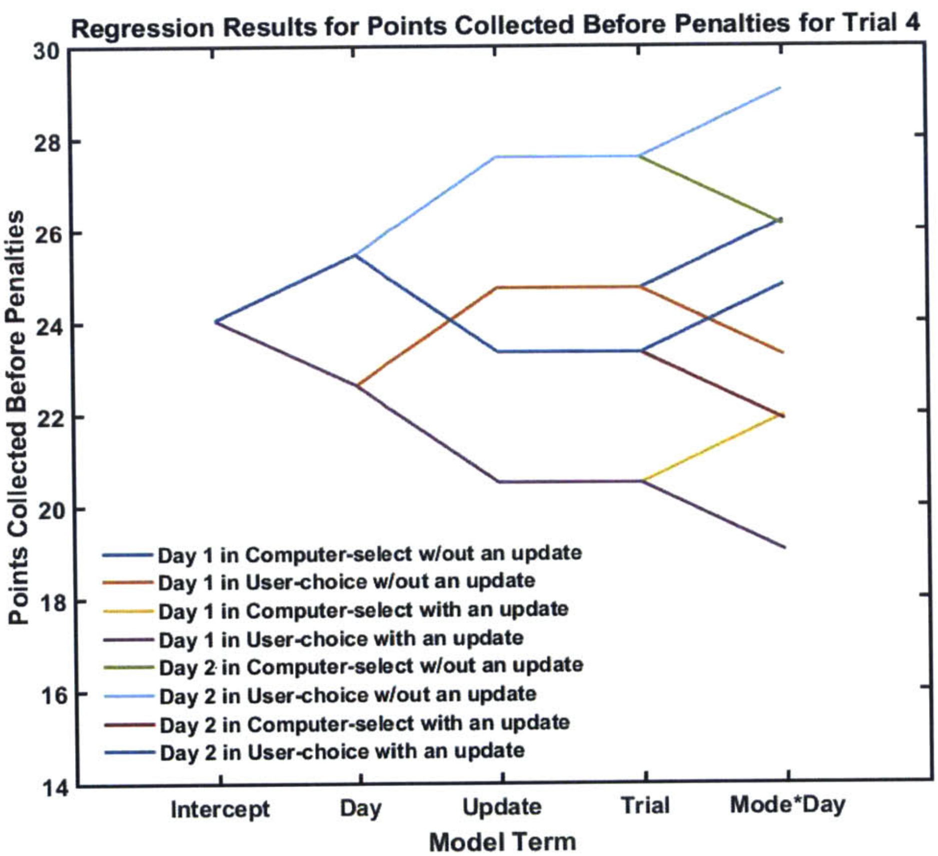


Figure Q-4: Impacts of the Individual Elements on the Linear Regression Model for Points Collected before Penalties for Trial 4

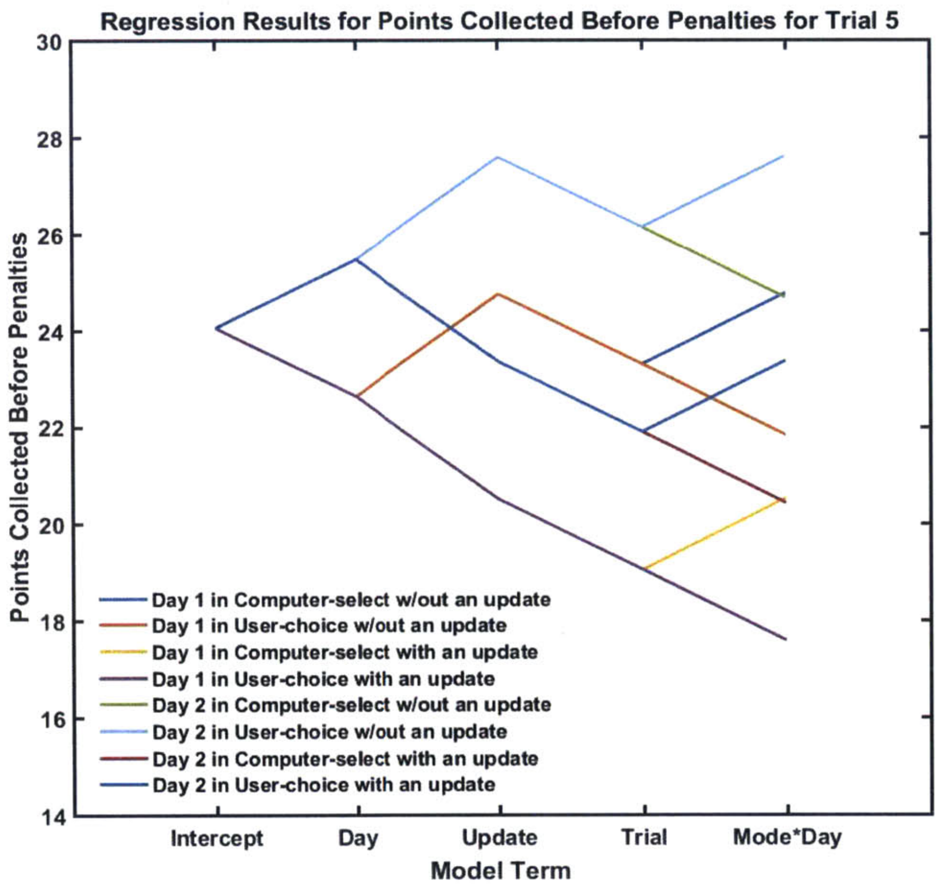


Figure Q-5: Impacts of the Individual Elements on the Linear Regression Model for Points Collected before Penalties for Trial 5

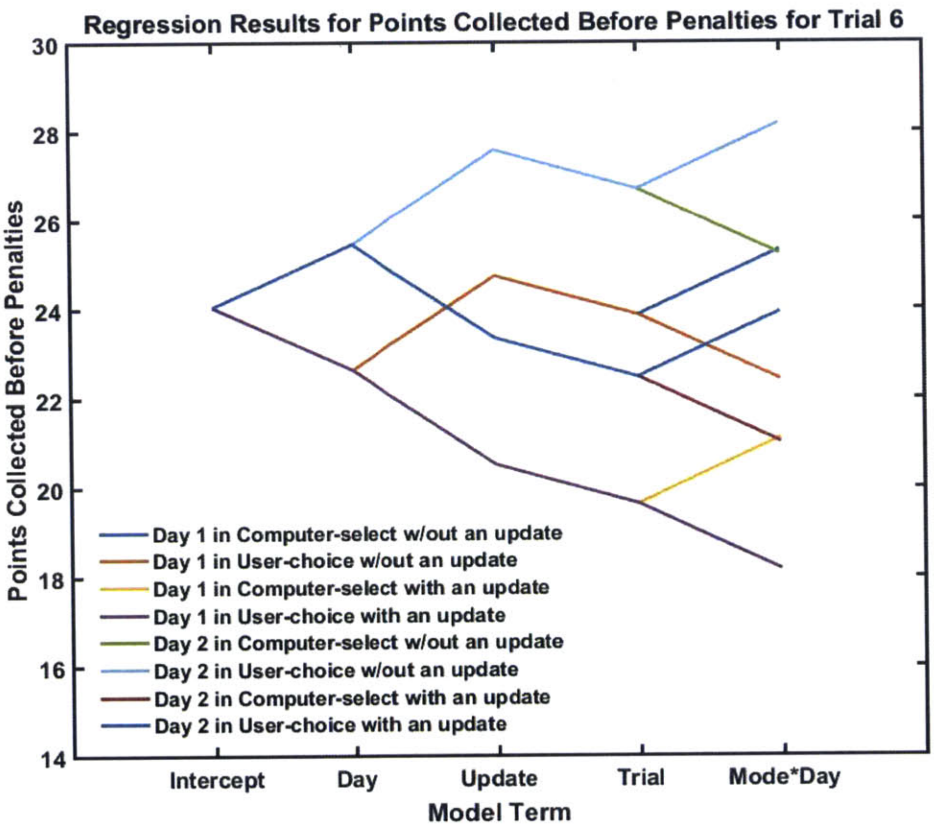


Figure Q-6: Impacts of the Individual Elements on the Linear Regression Model for Points Collected before Penalties for Trial 6

Table Q.2: Delta 1st Route Points Before Penalties

Variable	$\hat{\beta}$	SE	Z	p
Intercept	2.10	1.34	1.57	0.116
Day	-1.54	0.36	-4.27	<0.001
Update	2.18	0.36	6.11	<0.001
Order	2.13	1.08	1.97	0.049
Mode	3.42	0.38	9.09	<0.001
Plan Time	0.06	0.02	3.51	0.002
Mode*Day	1.51	0.36	4.22	<0.001
Mode*Day*Update	0.90	0.36	2.52	0.012

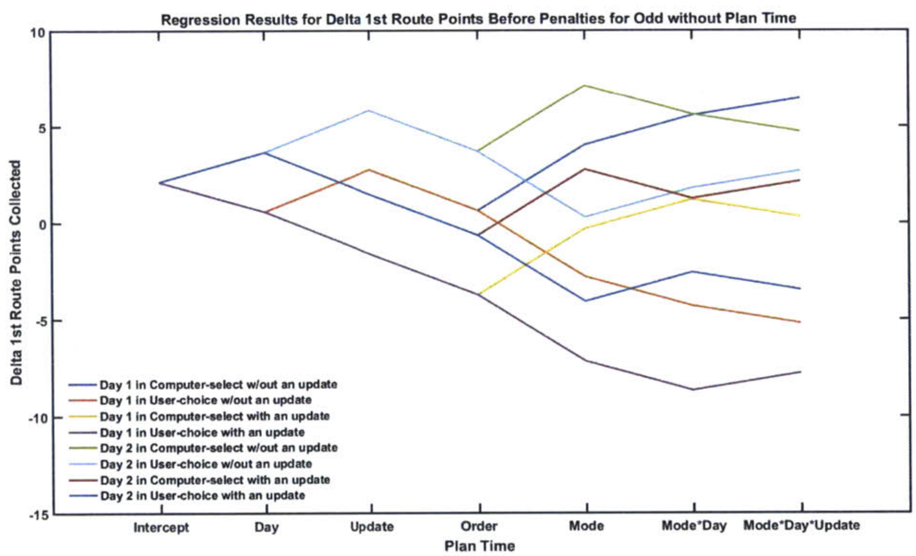


Figure Q-7: Impacts of the Individual Elements on the Linear Regression Model for Delta 1st Route Points Collected before Penalties for Odd Subjects

Q.2 Delta 1st Route Points Collected Before Penalties

Table Q.1 can be seen graphically in Figures Q-7 and Q-8 with the itemized impact of each model element without the inclusion of the covariate the plan time. The effect of the covariate on the model can be seen by second of plan time for each variation of mode, day and update in Figures Q-9 and Q-10.

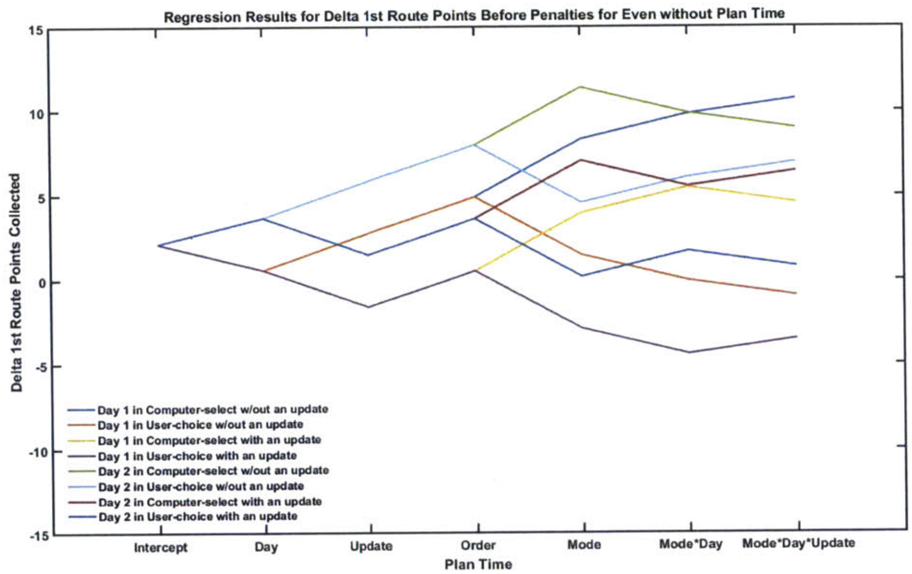


Figure Q-8: Impacts of the Individual Elements on the Linear Regression Model for Delta 1st Route Points Collected before Penalties for Even Subjects

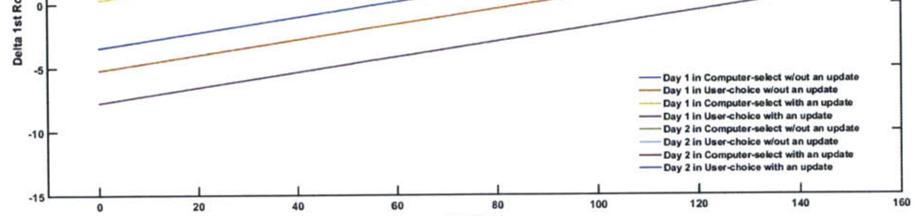


Figure Q-9: Impacts of the Individual Elements on the Linear Regression Model for Delta 1st Route Points Collected before Penalties for Odd Subjects with Plan Time

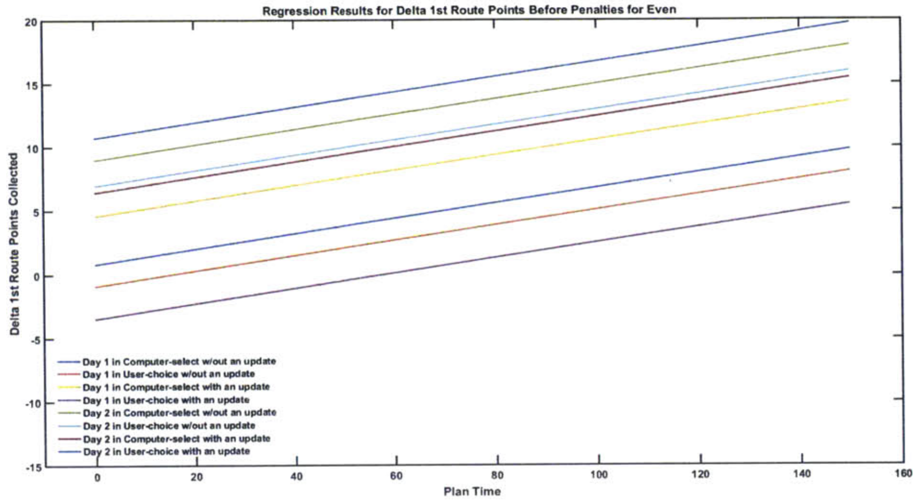


Figure Q-10: Impacts of the Individual Elements on the Linear Regression Model for Delta 1st Route Points Collected before Penalties for Even Subjects with Plan Time

Table Q.3: Total Points After Penalties

Variable	$\hat{\beta}$	SE	Z	p
Intercept	22.55	1.54	14.63	<0.001
Day	-1.95	0.44	-4.39	<0.001
Mode	0.99	0.45	2.23	0.025
Update	2.30	0.44	5.20	<0.001
Mode*Day	1.52	0.44	3.41	0.001

Q.3 Total Points Collected After Penalties

Table Q.3 can be seen graphically in Figure Q-11 with the itemized impact each additional term has on the estimate moving left to right.

Q.4 Delta 1st Routes Points Collected After Penalties

Table Q.4 can be seen graphically in Figure Q-12 with the itemized impact of each model element.

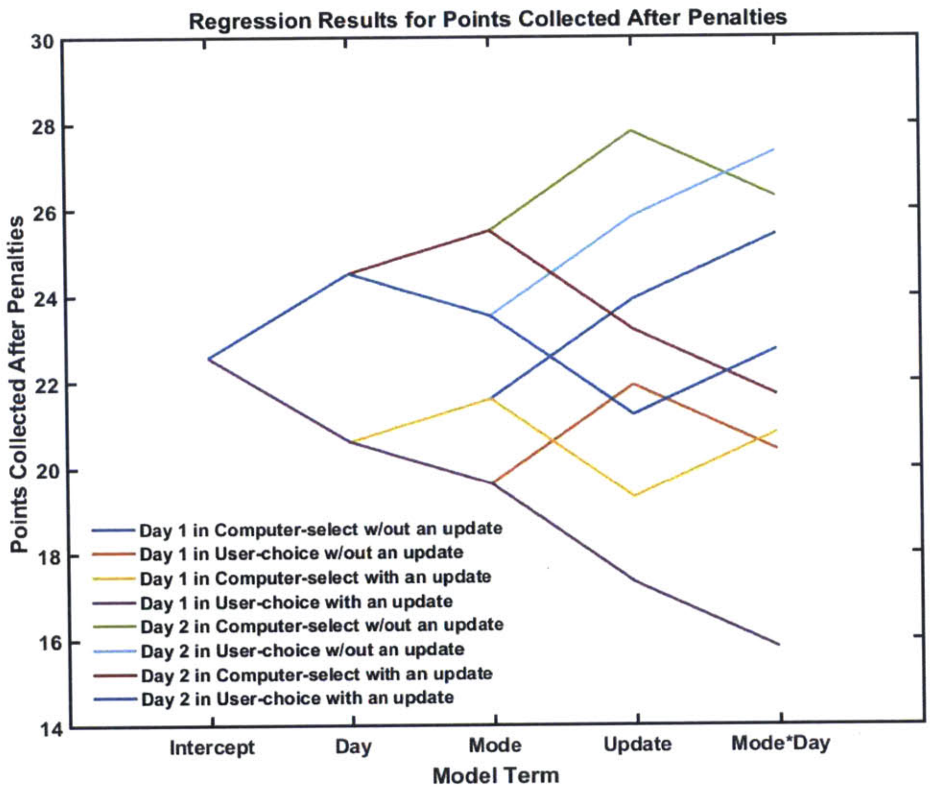


Figure Q-11: Impacts of the Individual Elements on the Linear Regression Model for Points Collected after Penalties

Table Q.4: Delta 1st Route Points After Penalties

Variable	$\hat{\beta}$	SE	Z	p
Intercept	3.18	1.59	2.00	0.046
Update	2.21	0.45	4.93	<0.001
Mode	3.86	0.45	8.57	<0.001
Day	-1.91	0.45	-4.25	<0.001
Mode*Day	1.48	0.45	3.28	0.001
Mode*Day*Update	1.20	0.45	2.67	0.008

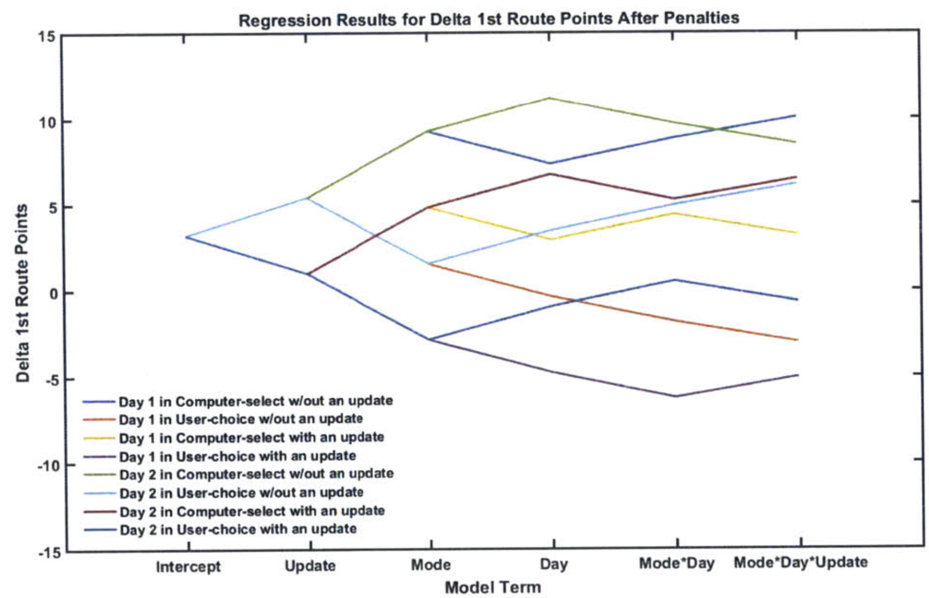


Figure Q-12: Impacts of the Individual Elements on the Linear Regression Model for Delta 1st Route Points Collected after Penalties

Appendix R

NASA TLX Raw and Weighted Components

Weighted TLX components were analyzed using Pearson Chi-squared tests. The mental component was shown to be dependent with day and order as seen in Figure R-1. While the performance component was shown to be dependent with order as shown in Figure R-2. The rest of the components were independent of mode, update, order, trial and day.

The weighted components are hard to extrapolate as meaning for the whole sample population, because they reflect user's interpretation of the divisions, which will be nulled out by the composite. However, these results show a general trend of the mental component having a greater impact on day 1 with it mostly accounted for in the odd group. The performance component shows that generally the odd group felt that performing well was a large part of their workload.

Raw TLX components were analyzed using Pearson Chi-squared tests. The temporal component was shown to be dependent with day and cross-effects between mode and day seen in Figure R-3. While the performance component was shown to be dependent with order as shown in Figure R-4. Finally the effort component was shown to be dependent upon order as displayed in Figure R-5. The rest of the components were independent of mode, update, order, trial and day.

As with the weighted components, the raw components do not tell a complete

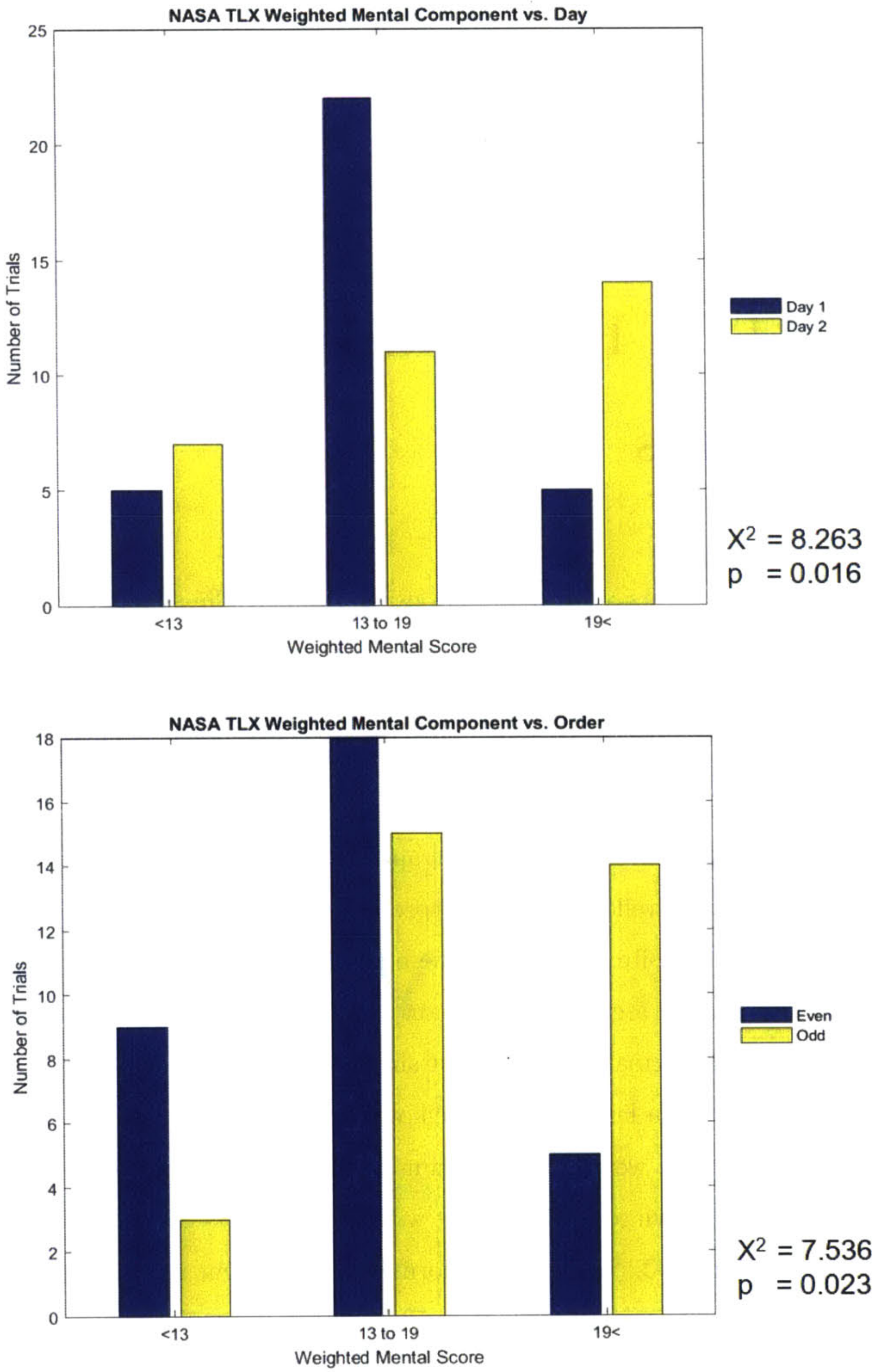


Figure R-1: TLX Weighted Mental Component Contingency Table Results

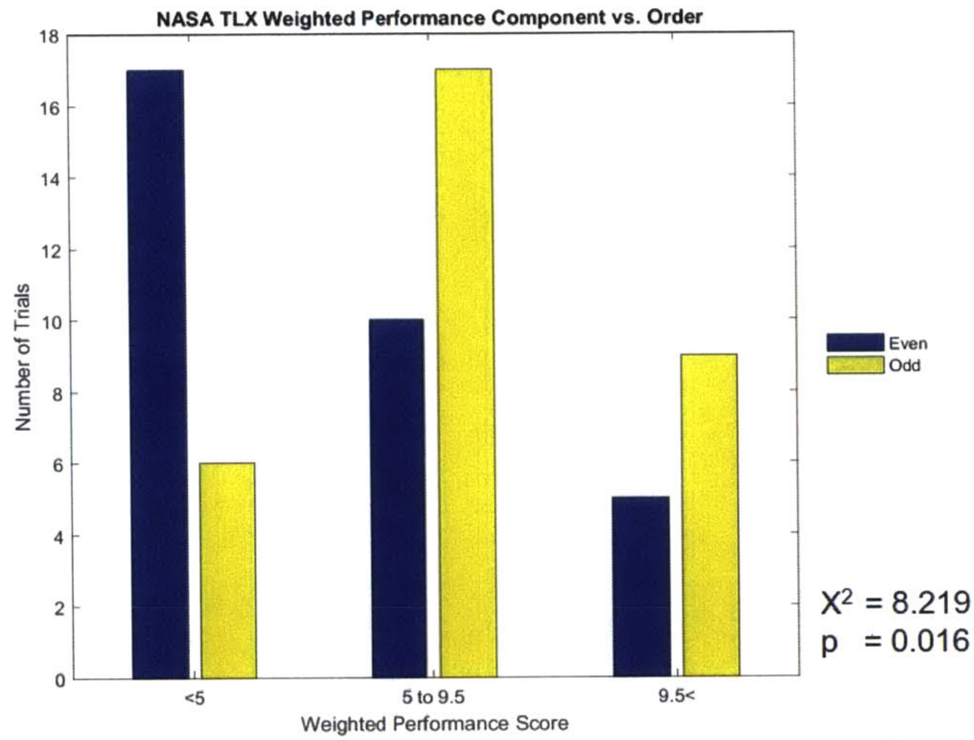


Figure R-2: TLX Weighted Performance Component Contingency Table Results

story, because of the interpretation of the meaning of each component. In this case the temporal component was generally higher on day 1, specifically with user-choice which reflects the time pressure of the task. The raw component of performance is consistent with the values shown in the weighted component with the odd group rating performance naturally higher. In addition, the even group felt that their effort was generally less.

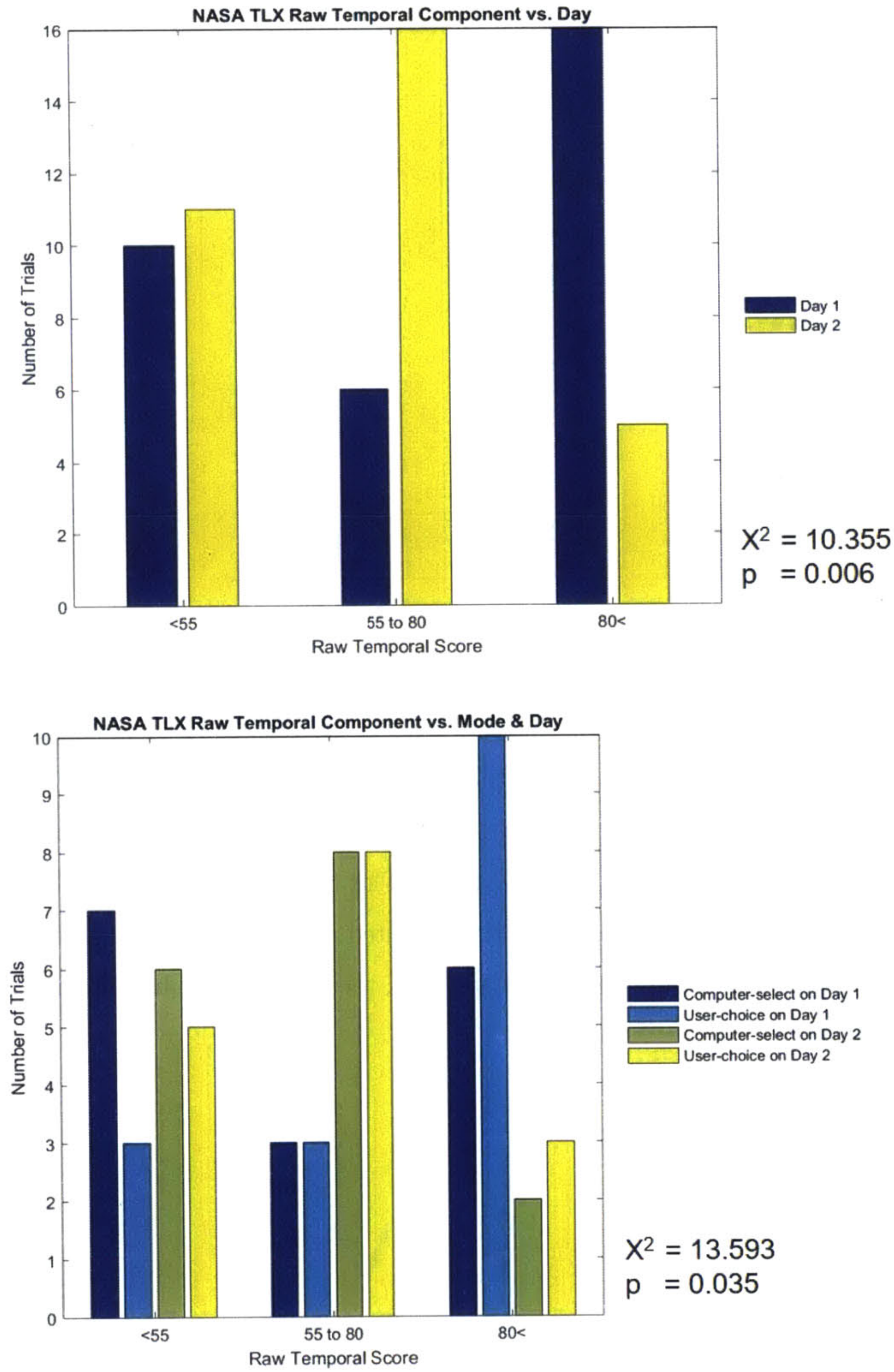


Figure R-3: TLX Raw Temporal Component Contingency Table Results

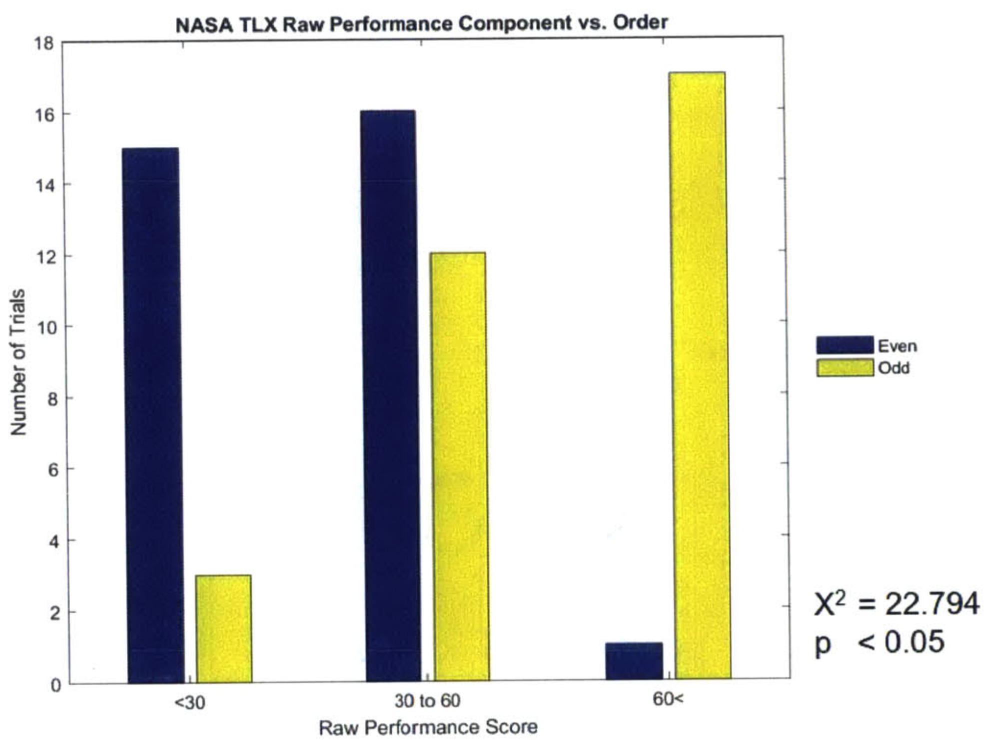


Figure R-4: TLX Raw Performance Component Contingency Table Results

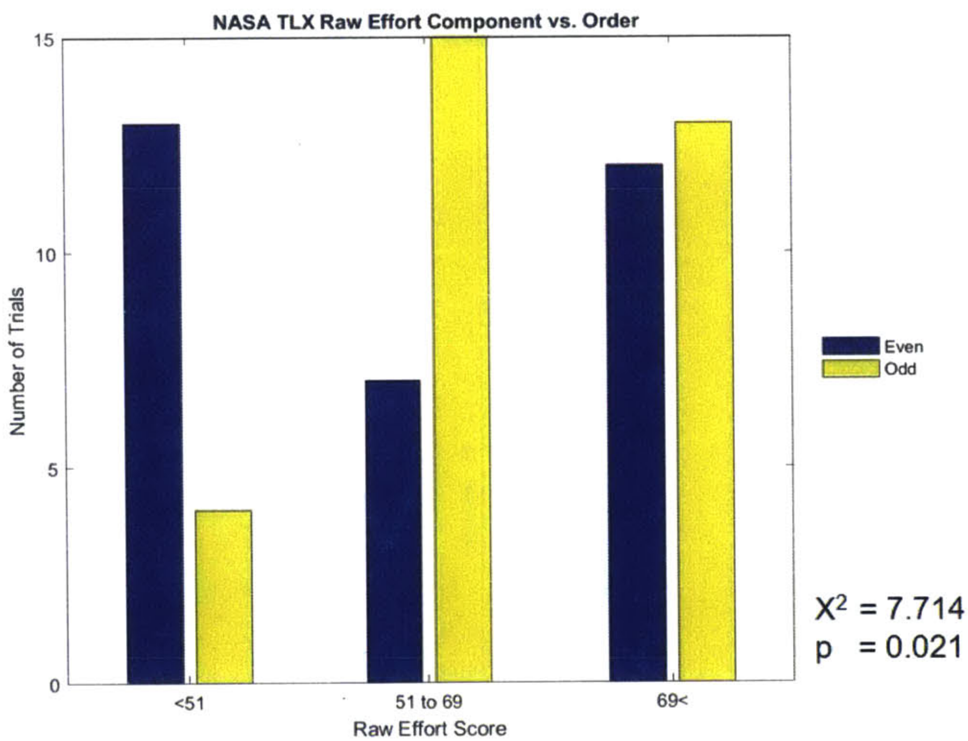


Figure R-5: TLX Raw Effort Component Contingency Table Results

Appendix S

Post-training Expectations for Most Challenging Aspect of Task

Below are the subject's expectations for the most challenging aspect of the experiment. The responses are grouped by similarity.

S.1 SA Task

- "situational awareness"
- "the situational awareness task"

S.2 Deviating from the App

- "Deciding if it's beneficial to go off the course to scan a qr code."
- "Deciding whether to listen to the app or not"

S.3 Updates

- "dealing with updates"
- "alerts mid-trial requiring New paths"

- “constant alerts"
- “updates in route"

Appendix T

Strategy Changes

Below is the compilation of all strategy responses grouped by similarity. Exact duplicates and empty responses were not included.

T.1 Aggressive

- "slightly more risky: going for extra points"
- "no. I was a little conservative and didn't reach for extra points since this was the first trial of the day."
- "slightly more aggressive"
- "I'm tired so I took the lazy route for this last trial"
- "move slower and get more points"
- "no work harder"
- "Need to be faster or take shorter route"
- "ignored an extra point, no time"
- "try and follow the fast route"
- "choose the fastest route"

- "go very fast, deviate when necessary, automation will correct"

T.2 SA

- "keep my head up more"
- "a little, now trying to look up more often"
- "trying to use the phone less and less"
- "look up more"
- "still trying to look up more and use map less"
- "look up more and don't follow a lot of the main track"
- "pay more attention to surroundings"
- "reviewing situational awareness info before taking survey"
- "disregard everything except points"

T.3 Update

- "trying to forecast alternate routes in case of floor closures"

T.4 Alerts

- "ignore path changes unless a floor is closed"
- "I ignore the warnings a little"
- "ignore the app route changes even if it says I have more time"
- "ignore all alerts and paths except floor closure"
- "no, I just press no to all route changes"

- ''no_ u may start accepting changes to make those Danny notifications go away"
- "don't use user choice feature, just use initial recommendation and maps"
- "check preview route that is most similar to the route I will take (I.e. including my own deviations) regardless of initial point and time estimate of app, and choose that one in hopes of having fewer updates."

T.5 Planning Route

- "completely followed the planned route this trial"
- "just use the initial path as a guideline"
- "need to look at point available initially"
- "stick to original path"
- "trust the initial path"
- "ignore route"
- "ignore everything except initial route"
- "try and follow the map more"
- "follow my own route"
- "try to decide the best route to take"
- "follow the original route"
- "follow the given path"
- "follow the generated path"
- "for automated mode I rely on my best predicted path and allow the autpmatip-nto catch up"

T.6 Willingness to deviate

- “starting to try not following the suggested riutes"
- “starting to not follow the initial path now"
- “yes, more willing to depart from the initial track"
- “followed app more"
- “followed app exactly"
- “Try and get more points"
- “deviating fromrelying on automated repath"
- “Yes, routes are more trustworthy. now I am to follow route but deviate where more than 4 points are immediately available"
- “app changes were more trustworthy, so I followed them"
- “Not since previous trials in comp choice mode. trusting app more in this mode"

T.7 Deviation Planning

- “try to get the bar codes worth the most evening if you take a completely arbitrary path"
- “follow a little of the given path and then deviate to get the 9 or 8 point bar-codes"
- “look up more and don’t follow a lot of the main track"
- “use the original path for guidance, but make my own path in the end"
- “use initial route as guideline, then change route to get the 8 or 9 point codes"
- “took detour for 9 points"

- “go rouge and add points if I think I can get them and they aren’t on the path"
- “yes - kinda going everywhere"
- “yes freelance"
- “Yes, i go with planned route but deviate greatly"
- “Staying with initial route, then deviating for higher point values on my own"
- “I don’t trust the app to account for 7+ point values properly in the route changes, so I do that myself"
- “deviate for 9 pointer"

T.8 Memorization

- “not yet, I may start memorizing the route"
- “do better pre memorization"
- “memorize and go"
- “yes, the app isn’t tracking as fast during the trials, so I will try to memorize more for the next runs"
- “remembered more from the map the first time"
- “memorize route at beginning, only change if floor closes. only using app to scan and double check point locations"

T.9 Time

- “no, I completely ignore the path other tha. to base how much time I should allot"
- “the app won’t always get you somewhere on time. always use best judgement"

- "just finish"
- "going for additional points that the path planner does not take me to if I believe I have time"
- "more conservative with time"

Appendix U

Best/Worst Parts of the App Responses

This appendix details subject responses about the best and worst parts of the app as taken following the use of each mode. The responses are grouped by similar responses.

U.1 Best Part of the App

U.1.1 Map

U.1.2 Side View

- "best was side view"
- "the tracking feature on the side view was the best"
- "side view showing point values"
- "Side view showing relative position of point values"
- "Side view showing point locations"
- "Side view showing point values"

U.1.3 General

- “best - map"
- “best - map"
- “best - the map"
- “best - map to follow"
- “it was good for navigation and figuring out where qr codes were"
- “best: two map views."
- “best: display of points on map"
- “best: map views"
- “best: map views"

U.1.4 Routes

U.1.5 Initial

- “the app gave good initial routes"
- “best was initial paths"
- “quality initial routes"
- “best initial path"
- “best was initial path"

U.1.6 Replanning

- “path projection to accumulate the most points was the best"
- “decent dynamic replanning"

- "the real time updates were the best"
- "best:automated updates."

U.1.7 Options

- "providing choices to the operator"
- "it allowed me to ignore possible changes"

U.1.8 Finish on time

- "the app made sure I got to the finish in time"

U.1.9 Workload Reduction

- "clear directions meant lowered mental workload"

U.1.10 WIFI Tracking

- "the up to date tracking was the best."
- "the tracking was the best."

U.2 Worst Part of the App

U.2.1 WIFI Tracking

- "the worst was getting lost when the app wasn't able to update the floor that I was on"
- "difficulties arose when the app was incorrect about which floor I was on"
- "worst - location tracking poor"
- "worst - location"

- “worst - lags a little and doesn’t give truly up to date info"

- “needs better timely location status’"

- “worst: occasional lagging"

U.2.2 User Involvement in Path Planning

Not Enough User Flexibility

- “perhaps letting the operator drag to adjust custom paths would be even more desirable"

- “inflexibility in route choices meant I had to settle for less points or be creative with which extra points I might go for."

- “it didn’t always choose the path that would lead to the most points,"

Too Much User Involvement

- “when a change came up all the new options were confusing to me . also having less human input and more information density could help"

- “automated New paths with less input from user"

- “the less input I give the better- faster scanning and better location determination is key"

- “needs less human input"

U.2.3 Alerts/Rerouting

Tracking

- “worst - annoying beeping cause I was off route time when the phone couldn’t track my position accurstely"

Frequency

- "it changed route way too often."
- "it made the new routes after an alert way too short"
- "the constant updates was the worst."
- "the co stand updates was the worst"
- "the updates were the worst"
- "worst: all the distracting notifications"
- "worst: frequency of notifications"

General

- "worst was updates"
- "worst was pop ups"
- "worst all updates"
- "worst was alerts"

Disregarding User Actions

- "sometimes rerouted despite me telling it not too"
- "worst: notified me too often, seemed to update even when I said no."

U.2.4 Map

- "the side view was the worst."

Appendix V

Mode Preference Responses

The below sections show the mode preferences of all subjects. Each subject consistently preferred one mode over the other. The responses show the reason for their preference. Duplicates and responses that just included the mode are not included.

V.1 Computer-select - 6 Subjects

- "fewer options at first and ability to clearly see some additional routes to the one given"
- "less options to pick at first and more freedom to go off track"
- "it interrupted me less and reduced workload"
- "less freezing"
- "it gave me more freedom to look around rather than having to manually change routes."
- "it allows be to look at my surroundings more."
- "less Input"
- "I didn't conform to the route much either way, and there were fewer notifications that way"

- "fewer notifications and less time wasted on them"

V.2 User-choice - 2 Subjects

- "it gave me a variety of paths to choose from and I could select the one that made the most sense to me (eg high points efficient path). however I ignored choosing between replans midtrial due to temporal demands"
- "it allowed me to choose routes that accrued more points while still reaching the goal in time. the extra workload of choosing the path ended up being preferable to the on-the-fly choices I was making to depart the computer chosen paths (in pursuit of more points)"
- "route changes could be ignored and it was easier to obtain more points"

Bibliography

[1] Sebastian Anthony. Think GPS is cool? IPS will blow your mind, 2012.

[2] Manish Bhojasia. Java program to implement traveling salesman problem using nearest neighbour algorithm.

[3] Meera M. Blattner, Denise A. Sumikawa, and Robert M. Greenberg. Earcons and Icons: Their Structure and Common Design Principles. *HumanâĂŞComputer Interaction*, 4(1):11–44, 1989.

[4] C. A. Bolstad, J. M. Riley, D. G. Jones, and M. R. Endsley. Using Goal Directed Task Analysis with Army Brigade Officer Teams. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 46(3):472–476, 2002.

[5] William Ching, Rue Jing Teh, Binghao Li, and Chris Rizos. Uniwide WiFi based positioning system. *International Symposium on Technology and Society, Proceedings*, pages 180–189, 2010.

[6] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. 2009.

[7] Nikolaus Correll. *Introduction to Autonomous Robots*, volume 53. 2016.

[8] Mica R. Endsley. A Survey of Situation Awareness Requirements in Air-to-Air Combat Fighters. *The International Journal of Aviation Psychology*, 3(2):157–168, 1993.

[9] Amos Freedy, Ewart DeVisser, Gershon Weltman, and Nicole Coeyman. Measurement of trust in human-robot collaboration. *Proceedings of the 2007 International Symposium on Collaborative Technologies and Systems, CTS*, pages 106–114, 2007.

[10] Stavros Garzonis, Simon Jones, T Jay, and Eamonn O'Neill. Auditory Icon and Earcon Mobile Service Notifications: Intuitiveness, Learnability, Memorability and Preference. *CHI 2009- Multimodal Mobile Interaction*, pages 1513–1522, 2009.

[11] Google. A new frontier for Google Maps: mapping the indoors, 2011.

[12] Peter E. Hart, Nils J. Nilsson, and Bertram Raphael. A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2):100–107, 1968.

[13] Sandra G. Hart and Lowell E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, pages 139–183. 1988.

[14] B a Huberman and N S Glance. Evolutionary games and computer simulations. *Proceedings of the National Academy of Sciences of the United States of America*, 90(16):7716–7718, 1993.

[15] Infsoft GmbH. Infsoft maps.

[16] Kim Jackson, Emily Vincent, Eric Jones, and Zahar Prasov. A Heuristic-Based Framework for Assessing Operator Trust in Autonomous Systems. In *NDIA Human Systems Conference*, 2015.

[17] Myounghoon Jeon and Yuanjing Sun. Design and Evaluation of Lyricons (Lyrics + Earcons) for Semantic and Aesthetic Improvements of Auditory Cues. In *The 20th International Conference on Auditory Display*, pages 6–7, 2014.

[18] R. Jia, M. Jin, and C. J. Spanos. SoundLoc: Acoustic Method for Indoor Localization without Infrastructure. *ArXiv e-prints*, July 2014.

[19] Jiun-Yin Jian, Ann M. Bisantz, and Colin G. Drury. Foundations for an Empirically Determined Scale of Trust in Automated System. *International Journal of Cognitive Ergonomics*, 4(1):53, 2000.

[20] Aaron W. Johnson. *Analyzing the Effects of Dynamic Task Allocation on Human-Automation System Performance.* PhD thesis, 2015.

[21] K Johnson, L Ren, J K Kuchar, and C M Oman. Interaction of Automation and Time Pressure in a Route Replanning Task. *International Conference on Human-Computer Interaction in Aeronautics (HCI-Aero)*, pages 132–137, 2002.

[22] David B Kaber, North Carolina, and Lawrence J Prinzel Iii. Adaptive and Adaptable Automation Design : A Critical Review of the Literature and Recommendations for Future Research. *NASA Technical Memorandum*, (September), 2006.

[23] Vikram Kaul, Christian Makaya, Subir Das, David Shur, and Sunil Samtani. On the adaptation of commercial smartphones to tactical environments. *Proceedings - IEEE Military Communications Conference MILCOM*, pages 2205–2210, 2011.

[24] B. Kidwell, G. L. Calhoun, H. A. Ruff, and R. Parasuraman. Adaptable and Adaptive Automation for Supervisory Control of Multiple Autonomous Vehicles. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 56(1):428–432, 2012.

[25] M Kozhevnikov and M Hegarty. A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory & Cognition*, 29(5):745–56, 2001.

[26] LadiesMan217. Indoor gps.

[27] J. D. Lee and Neville Moray. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10):1243–1270, 1992.

[28] J. D. Lee and K. A. See. Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1):50–80, 2004.

[29] Mohammed a H Lubbad, Mahmoud Z Alkurdi, and Aiman Abu Samra. Robust Indoor Wi-Fi Positioning System for Android- based Smartphone. *International Journal of Research in Business and Technology*, 3(2):159–162, 2013.

[30] Michael D. Matthews, Laura D. Strater, and Mica R. Endsley. Situation Awareness Requirements for Infantry Platoon Leaders. *Military Psychology*, 16(3):149–161, 2004.

[31] Christopher a. Miller. Trust in Adaptive Automation: The Role of Etiquette in Tuning Trust via Analogic and Affective Methods. *Proceedings of the 1st International Conference on Augmented Cognition*, page 559, 2005.

[32] Matthew Panzarino. What exactly WiFiSLAM is, and why Apple acquired it, 2013.

[33] Raja Parasuraman and Victor Riley. Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(2):230–253, 1997.

[34] Raja Parasuraman, Thomas B Sheridan, and Christopher D Wickens. A Model for Types and Levels of Human Interaction with Automation. *IEEE Transactions on Systems, Man, and Cybernetics*, 30(3):286–297, 2000.

[35] Amit Patel. A*, 2016.

[36] Nissanka Bodhi Priyantha, Hari Balakrishnan, Erik Demaine, and Seth Teller. Mobile-Assisted Localization in Wireless Sensor Networks. In *IEEE INFOCOM*, Miami, FL, March 2005.

[37] Yan Qi. An implementation of k-shortest path algorithm (java version).

[38] M. A. A. Rahman, Marzieh Dashti, and Jie Zhang. Floor Determination for Positioning in Multi-story Building. *IEEE*, 3:2576–2581, 2014.

[39] C. Raley, R. Stripling, a. Kruse, D. Schmorrow, and J. Patrey. Augmented Cognition Overview: Improving Information Intake under Stress. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48(10):1150–1154, 2004.

[40] Arie Rotstein, Omri Inbar, Tatiana Berginsky, and Yoav Meckel. Preferred transition speed between walking and running: Effects of training status. *Medicine and Science in Sports and Exercise*, 37(11):1864–1870, 2005.

[41] Tapio Saarelainen. Optimizing the Performance of a Dismounted Future Force Warrior by Means of Improved Situational Awareness. *International Journal on Advances in Telecommunications*, 5(1):42–54, 2012.

[42] SAILS Technology. Buildngo - indoor navigation.

[43] Juergen Sauer, Chung-Shan Kao, and David Wastell. A comparison of adaptive and adaptable automation under different levels of environmental stress. *Ergonomics*, 55(8):840–853, 2012.

[44] Adena Schutzberg. Ten Things You Need to Know About Indoor Positioning, 2013.

[45] T. B. Sheridan. A Scale of Degrees of Automation, and Where to Stop. In *Telerobotics, Automation, and Human Supervisory Control*, pages 357–358. 1992.

[46] Beom Ju Shin, Kwang Won Lee, Sun Ho Choi, Joo Yeon Kim, Woo Jin Lee, and Hyung Seok Kim. Indoor WiFi positioning system for Android-based smartphone. In *2010 International Conference on Information and Communication Technology Convergence, ICTC 2010*, pages 319–320, 2010.

[47] South Mobile. Crux indoor location.

[48] Mark St. John, Harvey S. Smallman, Daniel I. Manes, Bela a. Feher, and Jeffrey G. Morrison. Heuristic Automation for Decluttering Tactical Displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 47(3):509–525, 2005.

[49] svantetobias and Mwjuhl. a-star-java - default.

[50] Luke Swartz. *Why People Hate the Paperclip: Labels, Appearance, Behvavior and Social Responses to User Interface Agents.* PhD thesis, 2003.

[51] Megan Tschannen-Moran and Wayne K. Hoy. A Multidisciplinary Analysis of the Nature, Meaning and Measurement of Trust. *Review of Educational Research*, 70(4):547–593, 2000.

[52] Ananth Uggirala, Anand K. Gramopadhye, Brain J. Melloy, and Joe E. Toler. Measurement of trust in complex and dynamic systems using a quantitative approach. *International Journal of Industrial Ergonomics*, 34(3):175–186, 2004.

[53] Yolanda Vazquez-Alvarez, Ian Oakley, and Stephen A. Brewster. Auditory display design for exploration in mobile audio-augmented reality. *Personal and Ubiquitous Computing*, 16(8):987–999, 2012.

[54] Bruce N Walker, A Nance, and Jeffrey Lindsay. Spearcons: speech-based earcons improve navigation performance in auditory menus. In *Proceedings of the International Conference on Auditory Display*, pages 95–98, 2006.

[55] Lawrence R Wheeless and Janis Grotz. The measurment of trust and its relationship to self disclosure. *Trust and Self-Disclosure*, 3(3):250–257, 1977.

[56] B. G. Witmer, J. H. Bailey, B. W. Knerr, and K. C. Parsons. Virtual spaces and real world places: Transfer of route knowledge. *International Journal of Human-Computer Studies*, 45(4):413–428, 1996.

[57] Paul Woelfel. Wifi compass.

[58] By Ernest Wong. How Smartphones Will Reshape the Modern Battlefield, 2016.

[59] Jin Y. Yen. Finding the K Shortest Loopless Paths in a Network. *Management Science*, 17(11):712–716, 1971.

[60] David Zax. The Indoor Positioning System Era, 2012.