2017

Computer detection of spatial visualization in a location-based task

Georgi Iliev Batinov
Iowa State University

Computer detection of spatial visualization in a location-based task

by

Georgi Iliev Batinov

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Computer Science

Program of Study Committee:
Les Miller, Major Professor
Stephen Gilbert
Sunghyun Kang
Simanta Mitra
Johnny Wong

The student author and the program of study committee are solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2017

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I am grateful to Dr. Les Miller for his divine and untiring work.

I am also grateful to Dr. Stephen Gilbert, Dr. Sunghyun Kang, Dr. Simanta Mitra, Dr. Johnny Wong, Dr. Sarah Nusser, Dr. Maureen Kilkenny, Dr. Wallace Huffman, Dr. Dirk Reiners, Dr. Leigh Tesfatsion, Dr. L. Emil Kreider, Dr. Steven Huss-Lederman, Dr. Paul Campbell, Dr. Darrah Chavey, Dr. Frederick Anscombe, Dr. Matthew Fike, Dr. Walt Trahanovsky, Dr. Lucas Beverlin, and Dr. Adrian Silvescu for contributing in both intellectual and personal capacities to this project.

Thanks to Kofi Whitney for working shoulder-to-shoulder with me.

Thanks to Dr. Maya Taneva, my mother.

ABSTRACT

An untapped area of productivity gains hinges on automatic detection of user cognitive characteristics. One such characteristic, spatial visualization ability, relates to users' computer performance. In this dissertation, we describe a novel, behavior-based, spatial visualization detection technique. The technique does not depend on sensors or knowledge of the environment and can be adopted on generic computers. In a Census Bureau location-based address verification task, detection rates exceeded 80% and approached 90%.

CHAPTER 1. INTRODUCTION

1.1 General Problem

Through the proliferation of mobile devices, location-based software services have grown in both popularity and importance. According to Wilson (2012),

> "Location-based services (LBS)...provide functions that are location-aware, where the use of such services is predicated on knowledge of where the services are engaged. LBS are oft-referenced with regard to mobile devices, although LBS are not necessarily only used on mobiles."

A McKinsey Institute report estimated that in 2011, 28% of the U.S. population (87 million) used location-based services (Manyika et al. 2011). This massive user base requires varying levels of data fidelity. Accurate data is critical to organizations like the United States Census Bureau, which depends on authentic knowledge of every address in the nation to inform the distribution of $400 billion of federal monies each year (Census Bureau, 2015). The importance of location-based services extends to their attached user interfaces, which, from a software engineering perspective, are the loci of human-error management (cf. Maxion and Reeder 2005, p. 26).

Scientific and commercial interests have devoted considerable resources to interface research, but so far, the role of individual differences has been underrepresented. The literature suggests ample potential for improvement. For example, Benyon, Crerar, and Wilkinson (2001) derive the relevance of individual differences from a fundamental disparity between physical and digital artifacts. Information processing predominantly depends on symbol manipulation, so HCI systems are black boxes that can only be interrogated through their displays. By contrast, the user can employ multiple strategies to investigate physical artifacts. Therefore, cognitive differences may express more strongly on computer tasks (pp. 21-22).

Spatial visualization ability is one individual difference that has been associated with user performance. Ekstrom et al. (1976) defines it as “the ability to manipulate or transform the image of spatial patterns into other arrangements” (p. 173). Some known correlates of this aptitude are performance with command-line interfaces (Jennings, Benyon and Murray 1991, Benyon 1993), file system navigation (Vicente, Hayes, & Williges 1987), searching an information retrieval system (Downing, Moore, & Brown 2005), exploring a non-immersive virtual environment (Modjeska & Chignell 2003), web browsing (Zhang and Salvendy 2001), simulated driving (Andrews and Westerman 2012), and remote control of robots (Liu, Oman, Galvan, and Natapoff 2012).

The present work presents a spatial visualization detection technique drawing on behaviors on a location-based task, without external sensors.

1.2 Hypothesis

We hypothesize that on a location-based task that involves address verification, individual differences in spatial visualization ability lead to discernible variation in behaviors at the user interface, and that an algorithm can recognize the difference from the interface usage data alone.

This hypothesis will be validated if we observe algorithms detecting spatial visualization ability reliably enough to become viable in the real world. To accept the hypothesis beyond a reasonable doubt, we need to obtain favorable results persisting in multiple studies (two for the dissertation) using different software implementations, and in different environments. Based on results presented in the literature, a detection rate of 80% will have outdone prior research. This is because, other than in holistic cognitive fingerprinting (Chang et al. 2013), we have not seen any reports meeting a threshold of 80% correct detection of cognition-related variable. Kapoor, Burleson and Picard (2007), the classification apex, reported 79.17% accuracy in predicting user frustration. The prediction algorithm drew on a combination of software logs, posture-sensitive chair, galvanic skin response sensor, face tracker, and pressure-sensitive mouse. However, the authors' achievement in classifying transient emotions is in an altogether different research vein from the proposed dissertation's aim to detect a complex semi-permanent ability. Furthermore, external sensors are barriers to adoption.

1.3 Testbed

The application testbed is an address verification task performed by quality control officers of the U.S. Bureau of Census. The job of the Bureau is to collect and maintain statistics about the population and economy of the nation, with at least $400 billion of federal funds dependent on this information each year (Census Bureau 2015).

The Bureau of Census address verification task has the following desirable properties: it is a complex, professional, location-based task, and the Bureau's workforce is numerous and diverse.

The task consists of the following stages (stages 1 and 2 are interchangeable):

1. finding a specific address on a map;

2. locating the same address in reality;

3. ensuring the address is correctly reflected on the map and amending the map if necessary.

Bureau of Census survey takers are a diverse population and data fidelity is entirely dependent on employee competence. A computer device able to detect a user's suitability for the task during normal job duties opens a pathway to relevant adaptations that can be automated.

The dissertation will show connections between spatial visualization ability and user workflows. The detection technique will only employ user interface logs, because accounting for non-software behaviors requires specialized equipment (microphones, cameras, pressure sensors, galvanic skin response detectors, etc.) and would hinder adoption.

1.4 Contributions

The present work aims to bring four contributions to computer science.

1) Establish strong justification for greater incorporation of individual differences into the applied and theoretical research of intelligent interfaces. To our knowledge, there are few if any reports in the literature of detection of individual differences from professional tasks.

2) Demonstrate for the first time that it is possible for a generic computer device to recognize a cognitive ability. Literature reports of detection of other user variables frequently depend on external sensors and are therefore unsuitable for wide adoption.

3) Show that behavioral-based detection can circumvent the need to know what constitutes a user error, e.g. whether the address was correctly verified. Such a shortcut would be highly valuable in complex workflows, which are ubiquitous in professional computing, because the need for environmental information would be avoided altogether. As a result, both software and hardware designs can be simple without sacrificing the visualization detection capability.

4) Establish a relationship between spatial visualization ability and user preferences at the interface, with a goal to guide adaptive system design.

1.5 Organization

Chapter 2 surveys the existing literature and finds multiple sources that are almost relevant to the project, and a few that are directly related, due to the relative novelty of the behavioral approach. Chapter 3 discusses statistical outcomes from three human-subject experiments. Chapter 4 presents behavioral differences and infers decision models for the Paper Map experiment. Chapter 5 presents detection models and results. Chapter 6 investigates potential adaptations. Chapter 7 concludes the report.

CHAPTER 2. REVIEW OF LITERATURE

2.1 Introduction

User differences have always been of interest to system designers. Benyon, Crerar, and Wilkinson (2001) derive the relevance of individual differences from a fundamental disparity between physical and digital artifacts: information processing predominantly depends on symbol manipulation, so HCI systems are black boxes that can only be interrogated through their displays. By contrast, there are multiple strategies to investigate physical artifacts. Therefore, cognitive differences may express more strongly on computer tasks (pp. 21-22).

Spatial visualization ability is one individual difference that is frequently tested in experiments. Ekstrom et al. (1976) defined it as “the ability to manipulate or transform the image of spatial patterns into other arrangements” (p. 173). This aptitude correlates with performance in command-line interfaces (Jennings, Benyon & Murray 1991, Benyon 1993), file system navigation (Vicente, Hayes & Williges 1987), searching an information retrieval system (Downing, Moore & Brown 2005), web browsing (Zhang & Salvendy 2001), simulated driving (Andrews & Westerman 2012), and remote control of robots (Liu, Oman, Galvan, & Natapoff 2012).

Automated recognition of user variables is a large field that adjoins multiple disciplines, including computer science, psychology, human–computer interaction, ethnography, industrial design, many branches of engineering, instructional design and industrial ergonomics. There is an extensive list of reports on gathering information about the user, e.g. research on online learning environments like Blackboard and WebCT. But our variable of interest, spatial ability, has never been automatically recognized. In addition, intrinsic cognitive abilities in general are not represented as target variables. The most similar publications come from the field of adaptive interfaces.

Adaptive interfaces are encountered on any computer system that autonomously changes its interaction mode as a reaction to internal or external cues. Rothrock, Koubek, Fuchs, Haas and Salvendy (2002, pp. 58-63) use the term “variables calling for adaptation” and discuss user variables, situation variables, and system variables. User variables include an individual’s knowledge, performance, workload, personality and cognitive style.

Van Velsen, Van Der Geest, Klaassen, & Steehouder (2008) reasoned about “personalized” software as systems that employ some type of individual user model. With regard to usability, Van Velsen et al. claimed that “comparing a personalized system with one where the personalization has been removed is deemed a false comparison” (p. 265) based on statements in Höök (1997), Höök (2000), and Bohnenberger, Jameson, Kruger, & Butz (2002), in the sense that personalized systems have extra cross-sectional and longitudinal features which change the overall mix of utilities provided by the system. These are the words of just a few authors who express a zeitgeist of strong desire for personalization – which has rendered it a ubiquitous goal in most commercial and scientific software, and has invested it in multiple research domains.

Reports of user detection differ from our research in several ways:

1. To our knowledge, no attempts have been made to recognize spatial visualization ability or other specific cognitive abilities.[1] In contrast, we detect spatial visualization ability, which is linked to performance in many computer tasks. The detection is performed on a professional

[1] However, research exists on holistic cognitive fingerprinting, e.g. Chang et al. (2013)

task used by the Bureau of Census, and without knowledge of whether the user solved the task correctly.

2. Existing literature frequently uses external sensors. In contrast, our research detects spatial visualization ability based on ordinary user input. Detection is deployable on basic computing systems.
3. The published accuracy of recognition is relatively low, with correct classification of less than 80% of instances. In contrast, our research uncovered detection rates of 84% and 87%, which we deem practical enough for adoption in the real world.

2.2 Location of Research Objectives within the Scientific Field

User modeling for user interfaces is a topic within a broad area called “human-centered design”. Human-centered design refers to emphasizing user qualities during the software modeling process, as opposed to presenting an interaction protocol and demanding that users adopt it (Norman & Draper, 1986). While the field can be systematized in multiple ways, we will present Gleasure, Feller and O’Flaherty (2012)’s division of human-centered design approaches into four categories: metaphoric, idiomatic, contextualized, and foundational. Metaphoric approaches carry a real-world control (or other) convention over to the interface realm, attempting to gain usability through the familiarity of the metaphor (Gleasure, Feller, & O'Flaherty, 2012). An example would be a generic calculator program such as those shipped with most current operating systems.

Idiomatic design tries to co-opt operators’ knowledge of existing digital systems, perpetuating interaction modes that are already present in previously produced software. Unlike with the metaphoric approach, interface idioms (as well as linguistic idioms) have no comprehensible meaning outside of their intended use and therefore need to be explained

(Gleasure, Feller, & O'Flaherty, 2012). Idiomatic features include close buttons on most graphical interface windows and blinking cursors in command-line environments.

Contextualized design may be somewhat misleadingly named, as within Gleasure, Feller, and O’Flaherty’s classification it refers to “internal consistency within an application”, hence the term “contextualized”; but in practice, the methodology hinges on aligning the interface to user expectations by observing actual users. In this sense, the real context is the (sample) user base. Relevant investigative techniques span an array from in-depth ethnographic studies to iterative user evaluations (Gleasure, Feller, & O'Flaherty, 2012).

Research from the fourth category, foundational design, focuses on subconscious and unconscious factors in the interaction process, what the authors call “early perceptual and prejudicial aspects on interaction.” Foundational design incorporates findings from neurological and other sciences that may influence human behavior regardless of self-awareness. While one might raise the question whether foundational design is a part of contextualized design, Gleasure, Feller and O’Flaherty distinguish between the two based on how conscious user expectations are. Furthermore, foundational design does not necessarily demand user involvement at the interface creation stage, because readily available findings from relevant sciences, such as psychomotor studies’ outcomes, can be directly slotted into the process. Examples of potentially applicable results are known ergonomic concerns, for example, the difficulty with which the elderly notice some color combinations, or screen illumination levels that promote alertness (Gleasure, Feller, & O'Flaherty, 2012).

The metaphoric, idiomatic, contextualized and foundational design spaces form a continuum where user cognition becomes less and less conscious, with foundational design reaching into behaviors where awareness is irrelevant (Gleasure, Feller, & O'Flaherty, 2012). The present work elicited behaviors in the subconscious, foundational level.

Many of the papers in this literature survey depend on external sensors such as galvanic skin response sensors or gaze tracking devices. In contrast, our detection technique operates only on ordinary user input and is not computation-intensive, which enables deployment on most computer devices.

2.3 Literature Review Structure

Adaptation cues and responses naturally constitute a systematic description of adaptive interfaces. For our project, we are predominantly interested in cue acquisition subsystems. But the available literature groups along application domains rather than adaptation mechanics - a condition due to the interdisciplinary nature of the domain. In the rest of this chapter, we will present reports from neuroergonomics, educational data mining, personalized information retrieval, adaptive hypermedia, multimodal interfaces, accessible interfaces, task detection software, and industrial interfaces. None of the “comparable” experiments inform our approach to a significant degree due to the relative novelty of the research problem, so we assemble a context of research neighbors instead.

We encountered no research on address verification and cognitive abilities in a software engineering context. What follows will be a listing of research that neighbors ours mostly in the methodological area, but is otherwise of limited utility to the central question: having the software infer spatial ability as it observes the user.

2.4 Neuroergonomics

Parasuraman (2003) defined neuroergonomics as “the study of brain and behavior at work”. The field is an amalgamation of neuroscience and ergonomics, where ergonomics is “a scientific discipline concerned with the understanding of the interactions among humans and other elements of a system, in order to optimize human well-being and overall system performance" (Mehta & Parasuraman 2013). While such a postulation of the field should be generally applicable to the dissertation's goals, the current state of neuroergonomics reduces its practical relevance. In particular, the discipline is concerned with brain and body imaging through external sensors, and the focus is on understanding what happens physiologically within the user. Since physiological changes, such as neural activations, can refer to particular cognitive states in considerable detail, some form of quasi-mind-reading appears to be a long-term goal. At the same time, the necessary equipment for physiological detection is rare and expensive, and therefore unsuitable for the professional tasks we targeted. For example, Sciarini, Grubb, & Fatolitis used an electroencephalograph to examine workload changes on a Stroop task, where a word for the name of a color is presented in a different color to induce cognitive dissonance. The authors detected higher workload when the named color and the actual color were mismatched, but their results depended on the presence of an expensive external device and did not include automatic prediction. Similarly, Sciarini, Fidopiastis, & Nicholson (2009) were able to associate inter-beat intervals of the heart to spatial ability during a Tetris-like task, but attempting to replicate their results would require an electrocardiograph to be attached to a participant – a condition unfavorable to our goal of using generic computer devices for prediction.

Reeves et al. (2007); and Reeves & Schmorrow (2007) survey older adaptive systems triggered by physiological signals.

2.5 Educational Data Mining

The field of educational data mining frequently includes user modeling. Romero & Ventura (2010) define educational data mining as “an emerging interdisciplinary research area that deals with the development of methods to explore data originating in an educational context. EDM uses computational approaches to analyze educational data in order to study educational questions” (p. 1). The major sub-fields are educational hypermedia and intelligent tutoring. Calvet Liñán & Juan Pérez (2015) distinguish educational data mining from learning analytics along several dimensions. The most salient difference is that learning analytics is primarily concerned with empowering human decisions and strategic involvement, while educational data mining places an emphasis on automated discovery and adaptation (pp. 105-106). We next present several research reports from the field that are related to our project.

Antonenko, Toy, & Niederhauser (2012) reported two cases of student workflow differentiation based on cluster analysis of server logs of an online learning environment. Their research follows a user modeling pattern that persists throughout the field of educational data mining: decision models describe states of learning, a highly mutable variable, which makes them only marginally useful for our purpose, which is to investigate a cognitive competence that is immutable in the short term. In the first experiment, education students were asked to assume the role of high-school teachers and write a report recommending solutions to a school incident. Based on server logs of time spent visiting relevant resources, irrelevant resources, and writing, participants were grouped into “discriminating investigators”, “non-discriminating investigators”, “non-discriminating writers” and “writers”. Investigators spent more time visiting resources than writers, and discriminating participants devoted less time to irrelevant resources. Clustering identified non-discriminating investigators as having an inferior strategy.

In the second experiment, collaborating groups of three or four students had to select a mortgage plan given a complex list of requirements. Cluster analysis pointed at high-performers spending more time working on tailoring the problem submission and progressing at a steady pace, while low-performers spent more time visiting the available information resources and also started working late. In this report, we see some strategy differentiation between high and low performers, a workflow outcome that is crucial to the success of this dissertation. What is less useful to our research effort is that the strategy differentiation is an isolated observation not connected to cognitive aptitudes.

Other recent examples in hypermedia-based educational data mining include Del Puerto Paule-Ruiz, Riestra-Gonzalez, Sánchez-Santillan, & Pérez-Pérez (2015), who mined six association rules from hypermedia logs that inferred whether a student would pass or fail a course with greater than 97% accuracy. Xing, Guo, Petakovic, & Goggins (2015) compared the performance of genetic algorithms, Naïve Bayes, and several other machine-learning algorithms to predict final grades in an online mathematics course. Campagni, Merlini, Sprugnoli, & Verri. (2015) used clustering and sequential pattern algorithms to infer that college students who kept close to the ideal sequence of computer science exams during their college career graduated faster and with higher grades. These reports are interesting in associating behaviors with final outcomes, but they have a limit in their utility to the dissertation work, because the target variable, learning performance, is highly volatile.

The other part of educational data mining, Intelligent Tutoring Systems, emphasizes workflow analysis and has an ongoing interest in user modeling. Older papers that inform the background of our research include Kinnebrew & Biswas (2012), who identified frequently occurring online reading patterns for low-performers and high-performers on a climate change study topic. The source data was the sequence of links visited and the time taken with each link. High-performers tended to re-read important pages and were productive in both long and short reading sessions, while low-performers were more successful in long reading sessions and first-time reads. While interesting from a data mining perspective, the report by Kinnebrew and Biswas is representative of its field in being concerned primarily with learning performance, a target variable we already noted is volatile. The takeaways for our research from this neighborhood in the literature are mostly about recognition mechanics and, to a much lesser extent, about cognitive properties of human subjects..

A common type of user modeling in intelligent tutoring systems relies primarily on the correctness of student answers. Koedinger, McLaughlin & Stamper (2012) created models from large sets of student answers to automated tutors' questions. The answers were coded as correct and incorrect, and a student model consisted of a sequence of questions and the expected probability of failure on each one. The probability of failure was predicated on student proficiency, number of learning concepts involved, difficulty per concept, and experience with the concept. The authors data-mined the models to find superior concept combinations for the tutors. In a good student model, the probability of error was relatively stable and declining. Erratic jumps in failure rates or progressively increasing difficulty indicated a problematic teaching sequence. To maximize predictive power (root mean square error of the predicted sequence of correct-incorrect responses), the researchers regrouped concepts with a limited-brute-force method. New answer-concept combinations were formed by iteratively mutating existing models with portions of man-made models. Those models were created mostly independently by teachers and not expected to be remixed in a model search. The brute-force approach discovered combinations with higher predictive power than man-made models (which to that point were considered the standard), and more importantly, the improvements were localized to particular spots in the question sequences, pinpointing problematic teaching areas that had eluded human experts.

Waalkens, Aleven, & Taatgen (2013) tested three different approaches to tutoring single-variable linear equations. One tutor only allowed students to use the standard strategy for solving an equation, as taught in middle schools in the United States. Another tutor allowed students to use minor variations on the standard strategy, and a third accepted all possible solution paths. The authors found that allowing strategy variations improved learning, but did not find significant differences in teaching effectiveness between the multi-strategy tutor and the standard-strategy-with-variations tutor. This finding is relevant to our research in suggesting that the effort to accommodate multiple workflows in software pays off even for a highly constrained problem like single-variable equations.

Galán & Beal (2012) used EEG signals to predict student success on SAT-level mathematical problems. In this case, the cognitive model consisted of two brainwave functions denoting workload and engagement. The engagement signal predicted the first error in 80 percent of the cases based on the first 20 s of sensor data. The authors suggested adding a non-intrusive EEG module to intelligent tutoring sessions to help students stay interested in a problem. This research is relevant to the proposed dissertation in its methodology: employing machine learning techniques to infer a cognitive variable. But the invasive external sensors create a distance between the work and our project goal.

In a recent sensor-oriented report, Petersen, Pardos, Rau, Swigart, Gerber, & McKinsey (2015) predicted chemistry performance on an intelligent tutoring system from gaze tracking variables, with 66% accuracy. Their efforts showcase a drive in the community to improve tutor adaptivity by knowing more about the user, while the relatively low accuracy illustrates how challenging user inference is even with a sophisticated external sensor. However, their variables of interest, learning gains and problem-solving performance, do not directly relate to our variable, spatial visualization ability.

Argenta and Hale (2015) provide another example of the ongoing interest in inferring user state in intelligent tutoring systems. They reported automatically reordering learning modules within an educational game based on pre-test result and in-game scores, in order to maximize initial learning and subsequent retention. While the methodological story of how they connected user assessment with tutoring presentation is interesting to us, our research focuses on professional tasks in the real world which cannot be scored independently by the computer due to their open-ended nature.

The takeaway narrative from the educational data mining literature spanning thousands of articles is that there is high ongoing interest in inferring user states, and in particular in linking learning gains to user modeling. However, the variables of interest are not connected to our spatial visualization detection, thereby limiting the utility of educational data mining approaches to the methodologies involved.

2.6 Personalized Information Retrieval and Adaptive Hypermedia

User modeling is a foundational aspect in the domains of personalized information retrieval and adaptive hypermedia. These two domains have considerable overlap with educational data mining, which was covered in the previous subsection, and which constitutes a exceptionally large corpus of research reports. In this subsection, we will briefly cover applications outside of formal education. Our interest will be perfunctory due to the insufficient relevance of user models available in this space.

According to Steichen, Ashman, & Wade (2012), personalized information retrieval "typically aims to bias search results towards more personally relevant information by modifying traditional document ranking algorithms", while adaptive hypermedia biases "content retrieval and presentation by adapting towards multiple characteristics. Those characteristics, more typically called personalisation 'dimensions', include user goals or prior knowledge" (p. 1) In these two domains, personalization is achieved through content or result selection and is based on what topics were visited by the user. Steichen, Ashman, & Wade (2012) and Knutov, De Bra, & Pechenizkiy (2009) provide reviews of older adaptive hypermedia papers.

User modeling in the information retrieval domain, in one form or another, is often based on browsing or search histories augmented with rules or other structures, and the models themselves are information topic aggregations that are of interest to the user. A recent example from this research vein can be found in de Campos, Fernández-Luna, Huete, & Vicente-Lopez (2014), who express a popular view: "An accurate representation of the user profile is very important in order to obtain good retrieval results" (p. 1281). In their report, the authors modified a political document search engine algorithm to accommodate individual user models and serve more relevant results. Another recent study, Kotzyba, Siegert, Gossen, Wendemuth, & Nürnberger, (2015), investigated exploratory voice-controlled search specifically tailored for children in third and fourth graders. The drawback of the report was that it described a pilot-sized study with only five children tested. The user models were individual in nature and needed further research to be able to generalize outcomes.

Thomas, Bailey, Moffat, & Scholer (2015) estimated users' utility from search tasks, expressed as a user-desired number of relevant search results. The independent variables included search query length in characters, individual search word length, and several more complicated arithmetically-derived query-related characteristics, as well as a user's past search profile. Four factors diminish the applicability of this report to our investigation. First, the target variable, user utility from searching, is far removed from spatial visualization as a cognitive ability. Second, a user's utility of search results cannot be known with certainty even if the user reported a particular number of desired relevant results, as users themselves may not be aware of what their utility thresholds are. Third, the prediction performed relatively poorly against a baseline. Fourth, utilizing individual search profiles for each participant that were unrelated to the experiment of the study weakens the ability to predict a user's utility if these profiles were absent.

Brennan, Kelly, and Arguello (2014) investigated information retrieval tasks and associated higher spatial visualization ability with visiting and abandoning more search engine result pages, and with longer search queries. This is a particularly encouraging report, because it demonstrated behavioral differences between low- and high-spatial-visualization participants. We will go into more detail into this publication in Chapter 6, as it informs our understanding of the connection between spatial visualization and potential adaptations.

Overall, our impression from the domains of adaptive hypermedia and personalized information retrieval was that, on one hand, user models did not investigate variables applicable to our research, or, alternatively, if the user variables were relevant, the reports did not contain attempts at detection.

2.7 Multimodal Interfaces

The multimodal interfaces domain is concerned with human-computer interaction occurring through visual, aural and haptic channels. User modeling in this field can be elaborate due to the presence of multiple information streams from the variety of sensors and effectors.

Dumas, Lalanne, & Oviatt (2009) present an overview of older papers. The field overlaps with educational data mining, and some of the papers surveyed in that subsection of the literature review are relevant in the multimodal interfaces domain. For example, Petersen, Pardos, Rau, Swigart, Gerber, & McKinsey (2015) predicted user learning from an intelligent tutoring system from electroencephalograph feeds, while Galan and Beal (2012) used an electroencephalograph to predict success on SAT-level mathematical problems.

User modeling in multimodal interfaces is interesting to us due to potentially suggesting approaches to harvesting and processing data for automated inference of spatial ability. However, to the extent we have surveyed the literature, we have not encountered a report that directly informs that goal. Instead, recent examples in the field classified student dialogue utterances based on gestures and postures (Ezen-Can, Grafsgaard, Lester, & Boyer 2015), predicted learning style on a basic mechanical engineering task from speech, gesture and electro-dermal sensors (Worsley & Blikstein 2015), predicted user choice of graphical or voice interface (Schaffer, Schleicher, & Möller 2015), or predicted mind wandering while reading electronic text from gaze tracking (Bixler & D'Mello 2015). Worsley & Blikstein (2015) presented more elaborate user models, but used external sensor instrumentation and did not relate to spatial visualization. Schaffer, Schleicher, & Möller (2015) used a generalized utility user model to predict what they suspect are individually differentiated users again with the help of external sensors, and their variable of interest is not pertinent to our research. Bixler & D'Mello (2015) are representative of a large and long-running gaze-tracking research direction that has had mediocre success in predictive accuracy, with this instance reporting 72% correct classification over a baseline of 60%. Overall, the multimodal interfaces domain is a potential source of technique inspiration in user modeling, but does not inform our research substantially.

2.8 Accessible Interfaces

The domain of accessible interfaces and assistive technologies, however, does provide previous research that is relevant to our work. There are documented efforts to create adaptive applications to furnish personalized aid based on user interactions alone. Taylor, Sr., et al. (2009) were able to automatically modify the appearance of a web page presented to older adults in order to minimize errors on a web use task. The adaptation was triggered by interaction errors: mouse-click errors, scrolling errors, and content access errors. The resulting system provided performance that was not significantly different from the performance of a system where a psychologist had determined the interface customizations for the users. Both the adaptive system and the psychologist-determined system exhibited considerably better performance than the baseline, non-adaptive system. Unlike Taylor, Sr., et al. (2009), we focused on a substantially complicated map survey task. Furthermore, we did not tackle accessibility challenges, but rather more demanding workflow differentials among physiologically capable users.

Another relevant effort with the goal of improving accessibility for older adults is Hourcade et al. (2010). The authors' system, PointAssist, detected mouse-pointing errors exhibited by the elderly and selectively turned on pointer slowdown to assist the user with hitting the interface target. Again, our goal is not accessibility, but support for physiologically capable users. Additionally, our research pursues a more complex workflow efficiency improvement.

Gonzalez-Rodriguez et al. (2009) introduced GADEA, an interface personalization system which employed a mixture of adaptive and adaptable behavior. The system aimed to improve accessibility, and personalized user-facing dialogs for ability differentials like typing speed and vision accuracy. Only dialogs were monitored and adapted, with fuzzy logic compounding about fifty rules to reach final layout decisions. Adaptability in GADEA depended on questions about age, disabilities, and personal preferences, asked in the beginning of the interactive session. Additionally, for adaptivity, GADEA included background monitors that tracked user behavior at a dialog. They measured typing speed, pointing speed, mouse motion accuracy, user reaction time and others. Numeric readings of these variables were converted to categories, e.g. "low visual precision", and used in the fuzzy logic aggregator. Its output was categorical and could be converted back to percentage values for scaling visual objects. An example adaptation rule was:

IF USER_MOVEMENT_PRECISION IS LOW

AND USER_VISUAL_PRECISION IS HIGH THEN

INTERACTIVE_OBJECT_SIZE IS BIG.

The authors piloted the system with 26 participants divided into five groups according to their visual accuracy. GADEA created dialogs specific to each group for five separate messages, resulting in 25 dialogs in total. Participants were asked to indicate their preferred dialog out of the five tailored choices for each message. The authors reported percentage of participant preferences that matched GADEA's suggested personalization.

There are multiple methodology leads in the GADEA framework, but it did not infer user cognitive ability and took a "dragnet" approach to usability, which is quite useful for specific applications such as critical systems monitoring.

2.9 Task Detection Software

Task detection is a domain tangentially relevant to our research in attempting to infer user workflows in advanced environments. Two example older task detection publications are Rath, Devaurs, & Lindstaedt (2010) and Rath (2010). The authors used machine learning techniques to classify interactive tasks on desktop computers. The interactive tasks were complex and spanned multiple applications. The authors constructed classification features from document content, application identifiers (e.g. Microsoft Word), window names, user actions, intra-application interface tracking (through Microsoft's accessibility framework), and users' application-switching patterns. Interface events and interface components were highly ranked as classification attributes. A recent report on task detection appeared in Mirza, Chen, Hussain, Majid, & Chen (2015), where the authors attempted to discriminate between desktop activities during multitasking. While there is superficial likeness between papers in this area and our research - the user modeling features include graphical user interface events - the goals in the field are completely dissimilar from our direction of investigation, and therefore the utility of the literature is limited to possibly intriguing data-mining techniques.

2.10 Industrial Interfaces

User modeling application domains presents itself in systems supporting industrial and military operators: fighter pilots, industrial process attendants, air traffic control personnel and others. There is a sizable collection of publications detailing context-aware interfaces for industrial and military workflows, but the presented systems do not rely on user characteristics alone. In almost all cases, there are other environmental sensors that inform the software. In contrast, the proposed dissertation will rely only on ordinary user input to make decisions.

In this domain, an older publication that only considered user characteristics is Yen & Acay (2009). Their system changed the user interface to an air traffic control task based on detected user errors, completion time, and number of user actions, complemented by mutations introduced by a genetic algorithm. The adaptation and evaluation process was sequential and iterative:

1. users completed the task on one interface variant,

2. performance was used to rank the current interface against all previously tested variants,

3. a set of new interfaces was generated via the genetic algorithm, and

4. the process was repeated 80 times.

In case the genetic algorithm produced multiple interfaces in a single generation, the authors would only present a single variant for human user evaluation, discarding the rest by extrapolating performance based on the observed mean and variance from historic evaluations. The underlying assumption for the extrapolation appears to be that user performance has a Gaussian distribution with the historically observed mean and variance. The utility of Yen & Acay's work to our dissertation project is limited, because their framework made it possible to know when the user committed an error.

A relatively recent interface project with industrial implications is presented in Chang et al. (2013). The authors developed a personal keystroke authentication system based on inter-keystroke timings during typing tasks. The software was aware of the individual cognitive idiosyncrasies of its users exhibited in their inter-keystroke delays, being able to correctly identify the user with precision exceeding 98%. Since the research project is aimed at an overall cognitive "fingerprint" of a user, we cannot directly connect the outcomes to our research. However, the sequential and timing user input features used in the recognition task inform the machine-learning methodology we intend to use for the dissertation project.

2.11 Conclusion

In this chapter, we touched on an assembly of fields that contain publications which are "research neighbors" to the present work in predominantly methodology. More generally, we have not been able to discover reports investigating persistent cognitive abilities within a software engineering context. This state of the literature may possibly be due to the scarcity of breakthroughs in classification with the variables that have been attempted so far, as well as to light interest in behaviors from an engineering viewpoint All of the surveyed fields seem to still be moving towards obtaining better results with their primary variables, which are unrelated to our project.

CHAPTER 3. DESIGN AND PERFORMANCE STATISTICS

This chapter describes three human-subject experiments investigating individual differences in address verification tasks. The address verification tasks under consideration consisted of the following stages (stages 1 and 2 are interchangeable):

1. find a specific address on a map;
2. locate the same address in reality;
3. ensure the address is correctly reflected on the map and amend the map if necessary.

The goals of this chapter are: (a) to convey the scope and intricacy of our data creation efforts; and (b) to present performance results which provide support for this dissertation's objectives, in the sense that recurring performance differentials may indicate the presence of systematic behavioral differences.

The Paper Map study will be described first. In it, participants verified addresses in the field with pen and paper. The study allowed us to observe between-user differentials in non-software address verification, and therefore establish the credibility of individual difference research in this area.

The second experiment in this chapter is the Stationary Simulation experiment, where participants verified addresses on a tablet device while sitting at a desk. Information about the address location in reality was presented in panoramic pedestrian-perspective photos of residential neighborhoods. The experiment allowed us to observe a fully controlled environment where both the available information about addresses and the verification workflow were constrained.

Lastly, in the Field-and-VR experiment, participants used a handheld device to verify addresses in both the field and a high-fidelity immersive virtual environment. This experiment allowed us to observe participants acting with considerable degrees of freedom.

3.1 Paper Map Study

Spatial visualization ability has been linked to performance on a variety of tasks. Some examples include command-line interfaces (Jennings, Benyon, & Murray 1991, Benyon 1993), file system navigation (Vicente, Hayes, & Williges 1987), searching an information retrieval system (Downing, Moore, & Brown 2005), web browsing (Zhang and Salvendy 2001), simulated driving (Andrews and Westerman 2012), and remote control of robots (Liu, Oman, Galvan, and Natapoff 2012). The existence of these prior reports supported the possibility of individual differences manifesting in address verification.

The design of the paper map study was based on a pen-and-paper protocol to avoid constraints associated with computers. As software and hardware could hamper the user with hidden workflow bottlenecks, removing both would allow the participant freedom of behavior. The literature suggested a second benefit to avoiding technology: if divergent behavior was observed on the core cognitive task, the differentials could be magnified in subsequent computer-based exercises (cf. Benyon, Crerar and Wilkinson (2001), pp. 21-22). A third advantage of a paper-only approach would be to support other computer experiments by providing a baseline of fundamental individual differences in address verification. The capability to compare statistical results from a paper-only study against results from a computer study would improve both plausibility and generalizability of inference.

3.1.1 Team roles

This research study was conducted in collaboration with Kofi Whitney, Les Miller, and Sarah Nusser. Drs. Miller and Nusser acted as faculty advisors, while Kofi Whitney and Georgi Batinov equally shared in the work of designing and executing the study.

3.1.2 Design

The experiment consisted of a cognitive test phase and a field exercise phase. Cognitive testing was performed in a room at the Iowa State University campus, while subsequent field activities occurred in a residential neighborhood of Ames, IA. In the field, twenty-six participants were asked to check whether seven addresses in a residential neighborhood were correctly reflected on a paper map, shown in Figure 3.1. Participants had to physically walk through the neighborhood, find the requested addresses, and amend the paper map if it did not accurately reflect reality. They were allowed to write and mark on the map as they saw fit.

3.1.2.1 Recruitment, compensation, and compliance

This human-subject experiment was approved after review by Iowa State University's Institutional Review Board. Participants were recruited through flyers posted on the Iowa State University campus, and public bulletin boards in grocery stores and churches in Ames, IA. Participants were also recruited through a posting on the computerized online Student job board maintained by Iowa State University. The compensation offered was a $10 Target Gift card for participating in the cognitive testing phase, and $20 for participating in the field phase. Completion of the phases was not necessary for compensation to be offered. Participants were apprised of their rights in the experiment through a standardized Informed Consent form (see APPENDIX B).

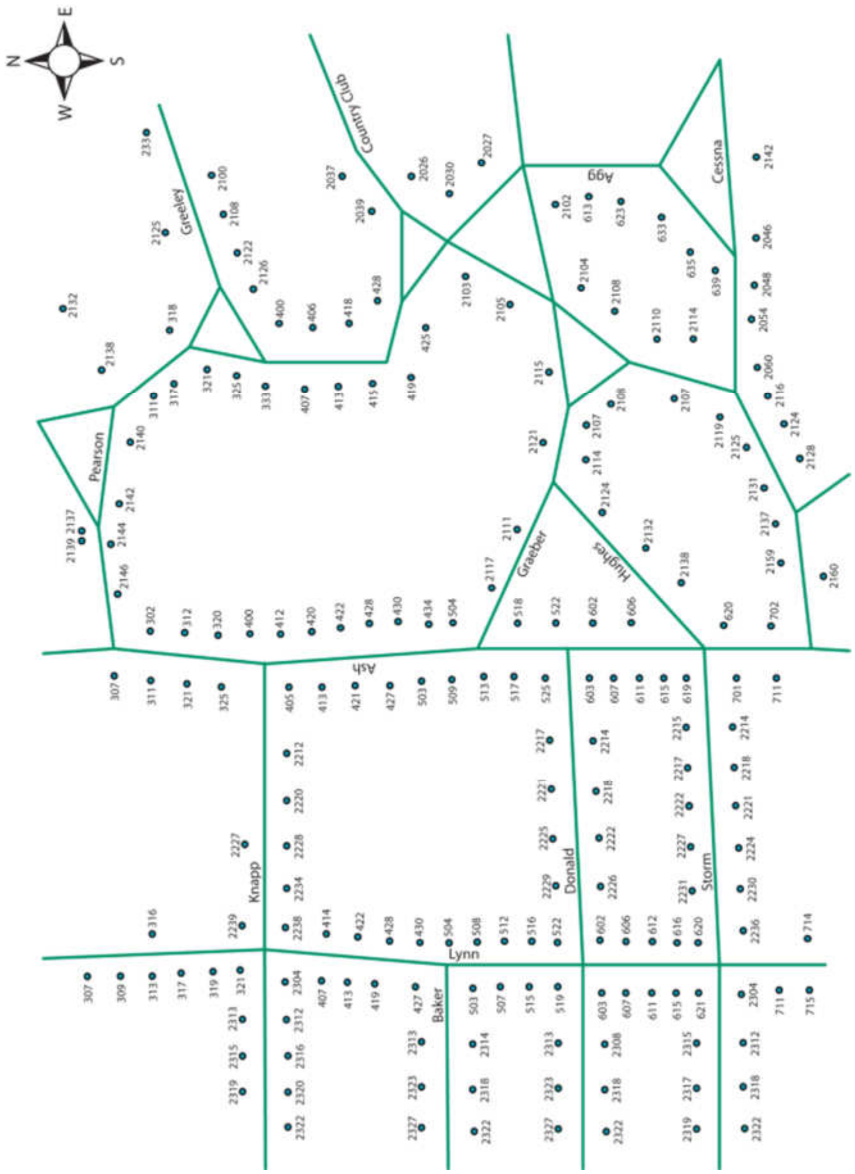


Figure 3.1 Field exercise map given to participants

3.1.2.2 Cognitive testing phase

During the cognitive testing phase, 99 participants were individually assessed on spatial visualization ability (VZ-2, Ekstrom et al. 1976), visual memory (MV-2, Ekstrom et al. 1976), and perspective-taking ability (Kozhevnikov et al. 2006) (see appendices E, F, and G). The location of the experiment was in an office on the Iowa State University campus. The cognitive testing phase lasted approximately an hour. At the beginning of the phase, participants had to read and sign an informed consent form. Tests were administered in immediate succession, with one-minute breaks between test sections and three-to-five-minute breaks between testsThirteen participants with spatial-visualization scores over or equal to 14.5 (out of possible 20) or perspective-taking scores over 29 were assigned to the high-spatial group. Thirteen other participants with spatial-visualization scores below 12 (out of 20) or perspective-taking scores below 11 were assigned to the low-spatial group. Perspective-taking scores have no defined maximum, but a score over 25 is considered high. Participants in the high-spatial and low-spatial groups were admitted to the field exercise.

The map contains highly irregular intersections and curving streets in the eastern half, while the western half contains right-angle intersections and straight-line streets. There were three addresses to verify in the “irregular” half of the map, three addresses to verify in the “ordinary” half, and one address to verify on the north-south street bisecting the map.

3.1.2.3 Field phase

During the field exercise, 26 participants (7 males) were taken individually to the exact same spot in a residential neighborhood in Ames, Iowa. They were trained on locating addresses in the field and the think-aloud protocol. An observer provided them with a clipboard with a paper map of the neighborhood on the front side (216x279 mm/8x11.5 inches, shown in Figure 3.1), a list of seven addresses taped to the back of the clipboard, and a four-colored ink pen. The next subsection reports on the details of target address selection.

The observer explained the task, the think-aloud protocol, and the possible results of each scenario. The goal of a participant was to determine whether the seven addresses were correctly reflected on the paper map. Participants would have to physically walk to an address in order to answer the question. If the map contained errors, they were expected to mark or write on the map to indicate the proper position of the address. Participants were further informed that experimenters were not interested in map errors that were not at the target addresses.

Figure 3.2 Example participant in the Paper Map Study. The paper map is affixed to the front of the clipboard, while a randomized list of target addresses identical to all participants is affixed to the back. The participant is holding a four-colored pen and is able to mark on both map and address list as desired.

Four outcomes were possible for each address: (1) *add-to-map*, (2) *move-on-map*, (3) *delete-from-map*, and (4) *confirm-on-map*. Participants were told to only work on the requested addresses and to ignore other possible errors on the map. Participants were not told that the map contained no errors outside of target addresses.

After the initial explanation, participants were asked to locate and verify three training addresses in the immediate vicinity on a simplified map with only two streets, while the observers answered procedure questions and provided feedback on the quality of the think-aloud.

At the end of the training session, observers answered any final questions participants may have had. They also explained that they would not talk or answer questions during the experiment, other than to prompt the participant to keep verbalizing or to ask about behavioral details. Observers (1) returned the participant to the exact location where all trainees started; (2) replaced the training map with the full exercise map; and (3) started an audio recorder (a Zoom H2 Portable Stereo Recorder, worn by the participant) and a GPS tracker (a HTC Android smart phone, carried by the observer). The GPS tracker was not given to the participant to avoid interrupting the workflow to time-stamp scenario completions. Additionally, observers walked behind the participant, establishing a close approximation of the exercise path. Observers wrote comments on standardized coding sheets of paper. A participant in the experiment appears in Figure 3.2.

After participants solved their final scenario, observers audio-recorded an exit questionnaire of 13 items detailing the participant's perceptions of the exercise (see APPENDIX I).

3.1.2.4 Map composition and target addresses

The study map (shown in Figure 3.1) contained two layers. Street layout and labeling were composed from the Census Bureau's TIGER/Line dataset, located on the Census Bureau website at https://www.census.gov/geo/maps-data/data/tiger-line.html. Address spots and labels were based on a set of parcel centroids furnished by the Story County Geographic Information Services Office. This is a governmental unit in Nevada, IA, USA, online at http://www.storycountyiowa.gov/index.aspx?NID=103. The address spots were moved on the map to align with buildings visible on geo-referenced satellite photos. The resulting map layout was an approximation of the Census Bureau's in-house visual presentation, which is unavailable to units outside the Bureau.

All participants verified the same seven addresses off of an identical randomized list order, and therefore could not benefit from inadvertent route hints. The list could be consulted at all times by flipping the clipboard. The opposite locations of the map and list meant that both could not be consulted at once unless the list or map were detached from the clipboard. This design choice made it obvious when users were checking the list of addresses. Participants were allowed to work on addresses in any order and could return to previously submitted scenarios as many times as they wanted. Only final answers were evaluated for correctness.

3.1.3 Results

The results in this section were published in Whitney, Batinov, Nusser, Miller, & Ashenfelter (2011). Table 3.1 presents the observed correlations between the three cognitive tests. The correlation values (*r*) are listed together with *p* values expressing the probability the correlation did not exist given the available test scores.

Table 3.1. Cognitive test score correlations.

COGNITIVE TEST	COGNITIVE TEST	n	r	p
Spatial Visualization	Visual Memory	26	0.54	0.00
Spatial Visualization	Perspective-taking	26	0.44	0.02
Perspective-taking	Visual Memory	26	0.36	0.07

The following performance ranges were observed in the study: 30 to 66 minutes for exercise completion times between 30 and 66 minutes, personal distances traveled between 1.10 mi and 1.88 mi (1.77 km and 3.02 km), and 0 to 3 incorrectly completed addresses per exercise.

Total time, *distance traveled*, *error pre-detection,* and the number of *errors made* by each participant were correlated with cognitive test scores. Table 3.2 presents the significant correlations between cognitive test scores and performance metrics. Spatial visualization test scores were negatively correlated with total time (*r = -0.44, p = 0.02*) and distance traveled (*n = 21, r = -0.65, p = 0.00*), revealing that lower-scoring participants tended to take longer and travel farther to complete the exercise. Perspective-taking scores were negatively correlated with total time (*r = -0.51, p = 0.01*), suggesting that, on average, participants with lower perspective-taking ability were slower in arriving at solutions. Additionally, both spatial visualization scores and perspective taking scores were positively correlated with error pre-detection, which tracks the tendency of participants to notice address errors while initially familiarizing themselves with the map. Correlations with pre-detection reveal that high-ability participants were more likely to detect flaws in the map model without needing cues from reality.

Table 3.2 Correlations of cognitive test scores and performance

COGNITIVE TEST	PERFORMANCE METRIC	n	r	p
Spatial Viz.	Total Time	26	-0.44	0.02
Spatial Viz.	Distance Traveled	21	-0.65	0.00
Spatial Viz.	Error Pre-detection	21	0.44	0.05
Persp. Taking	Total Time	26	-0.51	0.01
Persp. Taking	Error Pre-detection	25	0.49	0.01

3.1.4 Conclusion

Overall, the Paper Map study provided evidence that, for an address verification task, increased spatial visualization ability and perspective-taking ability correlate with better performance. The direction of the statistical connection was congruent with published findings in other exercise types (e.g. command-line interfaces (Jennings, Benyon and Murray (1991)), simulated driving (Andrews and Westerman 2012), remote control of robots (Liu, Oman, Galvan, and Natapoff (2012), suggesting that an address verification task is one more activity that is sensitive to spatial ability components, and spatial visualization ability in particular.

Evidence of performance differentials on the "baseline" task encouraged a search for behavioral differentials.

3.2 Stationary Simulation Study

This exercise investigated individual differences in address verification with significant constraints on both software workflow and information available in the panoramic views of the target addresses. Twenty-four participants used address verification software on a tablet device while viewing photos of a neighborhood. Participants were in a stationary seated position for the duration of the exercise, and the photos were displayed on two adjacent monitors (as shown in Figure 3.1). Participants had to amend address locations on the tablet to reflect the information presented on the monitors.

3.2.1 Team roles

This research study was conducted in collaboration with Michelle Rusch, Kofi Whitney, Les Miller, and Sarah Nusser. It was published as Rusch, Nusser, Miller, Batinov, & Whiney (2012). Drs. Miller and Nusser acted as faculty advisors, Michelle Rusch designed and executed the study, Georgi Batinov wrote the software, and assisted with the study design and execution. Kofi Whitney assisted with the study design, executed the study, and contributed to the software.

3.2.2 Design

The study consisted of two phases: a cognitive test phase and a computer exercise phase.

3.2.2.1 Recruitment, compensation, and compliance

Participants were recruited through flyers posted on the Iowa State University campus, and public bulletin boards in grocery stores and churches in Ames, IA. The compensation offered was a $10 Target Gift card for participating in the cognitive testing phase, and $20 for participating in the computer exercise phase. Completion of the phases was not necessary for compensation to be offered. Participants were apprised of their rights in the experiment through a standardized Informed Consent form (see APPENDIX K).

3.2.2.2 Cognitive test phase

In the cognitive examination portion of the experiment, participants had to solve three psychometric tests in the exact same sequence: Ekstrom et al.'s (1976) VZ-2 Paper-Folding test of spatial visualization ability and Kozhevnikov et al.'s (2006) Perspective-taking test from the Ekstrom et al. (1976) factor-referenced test battery (see Appendices E and G). The location of the experiment was in an office on the Iowa State University campus. The cognitive testing phase lasted approximately an hour. At the beginning of the phase, participants had to read and sign an informed consent form. Tests were administered in immediate succession, with one-minute breaks between test sections and three-to-five-minute breaks between tests. All tests by Ekstrom et al. were paper-based, while Kozhevnikov et al.'s Perspective-taking test was carried out on a desktop computer.

3.2.2.3 Computer exercise phase

For the computer exercise phase, twenty-four participants (twelve males) were taken individually to a room with the computer tablet and two adjacent twenty-inch LCD monitors shown in Figure 3.3. The location of the experiment was in an office on the Iowa State University campus. In a stationary seated position, participants used address verification software running on the tablet. Their task was to verify the map location of addresses in a town against photos of the addresses taken from a pedestrian perspective (Figure 3.3). The adjacent monitors showed a combined photographed view of two sides of the street at the target address. The observer explained the nature of the task and asked the participants to complete two untimed training scenarios, which were of similar type and difficulty as the experimental scenarios.

Participants were allowed to ask questions during the training. At the end of the training, the observer offered to answer any additional questions. Figure 3.5 shows a storyboard of the interface for one scenario.

Figure 3.3 Example participant in Stationary Simulation study (photo courtesy of Michelle Rusch). Participant is in a stationary sitting position at a desk, and the verification software is loaded on a tablet computer fixed in a stationary position. The two screens show photos of two sides of a street.

The remainder of this section presents an overview of scenario types and participant workflow. Further detail on the workflow pertains to software specifics and is described in the following "Materials" section.

In the experimental task, participants had to verify ten target addresses. Five scenario types were tested:

a) address needed to be added to the map;

b) address needed to be deleted from the map;

c) address needed to be moved to a different location;

d) address was present and required no corrective action; and

e) address was absent and required no corrective action.

Each scenario type was tested in two out of ten target addresses. Addresses and their sequence did not vary among participants. The software map and photos depicted Cedar Falls, Iowa.

The software for this exercise offered helper questions along the way. Figure 3.2 shows one path through the software, corresponding to scenario (b), “address needed to be deleted from the map”. Participants went through the following sequence while completing a target scenario:

locate the address on the photos;

1) answer a software question of whether the address exists;
2) find the address on the map, if possible;
3) answer a computer question of whether the address is on the map;
4) answer a software question of whether the address is in the correct location; and
5) add, delete, or move the address, if applicable.

<u>3.2.2.4</u> <u>Materials</u>

Georgi Batinov wrote the tablet software for the address verification exercise in the Java 1.5 programming language and the Swing graphical library. The software was loaded onto a Gateway M1300 tablet device with a 500 MHz CPU, 512 MB of random access memory, a 40 GB Hard Disk Drive and a 12.1-inch (307 mm) active matrix color screen with resolution of 1024x768 pixels (246 x 184 mm, see Figure 3.3). For the experiment, the tablet was oriented in landscape mode, with a horizontally positioned wide side of the screen. The dimensions of the software were smaller than the tablet display to more closely emulate the screen real estate of handheld device that could be used for address verification in the field. The interface area dimensions were 2 ¼ inches (57 mm) in width by 3 inches (76 mm) in height and the map area dimensions were 2 1/16 inches (52 mm) in width by 1 7/8 inches (48 mm) in height.

Kofi Whitney wrote the image display software in Java 1.6. Two copies with separate photo sets were loaded on two desktop computers driving 20-inch Dell LCD screens with 4:3 display ratios. The monitors were 16 inches (40.64 cm) wide and 12 inches (30.48 cm) tall. The verification software sent photo display commands to the desktop computers via a wired local area network. The adjacent monitors showed a combined photographed view of two sides of the street at the target address. Both desktops and the tablet ran the Windows XP Professional operating system. Participants used a stylus on the tablet's touch screen to perform software operations. Every time a participant signaled the start of a scenario, the tablet software commanded the display stations to change the environmental view. At the end of the two training scenarios, the displays were commanded to show red stop lights until the observer finished answering any last questions by the participant. Figure 3.3 displays the tablet computer and the environmental displays as they were used during the experiment.

There were two interface versions presented to participants: the "guided" interface and the "unguided" interface. The guided interface had additional elements compared to the unguided interface (Figure 3.4.). A yellow box at the top of the guided interface area contained all the steps necessary to complete the current scenario, with the current step highlighted. To the right of the yellow box, a white box contained an instruction on what to do for the current workflow step. The instructions changed as workflow steps progressed (Figure 3.5).

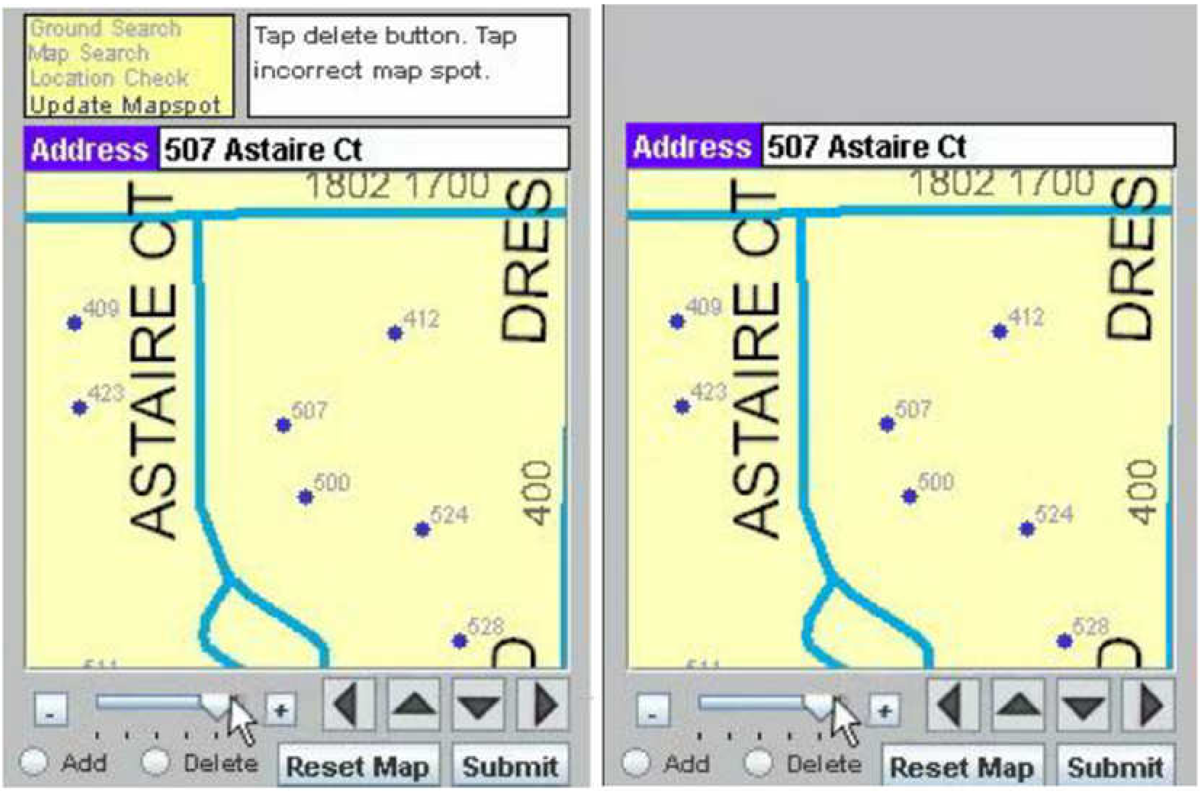
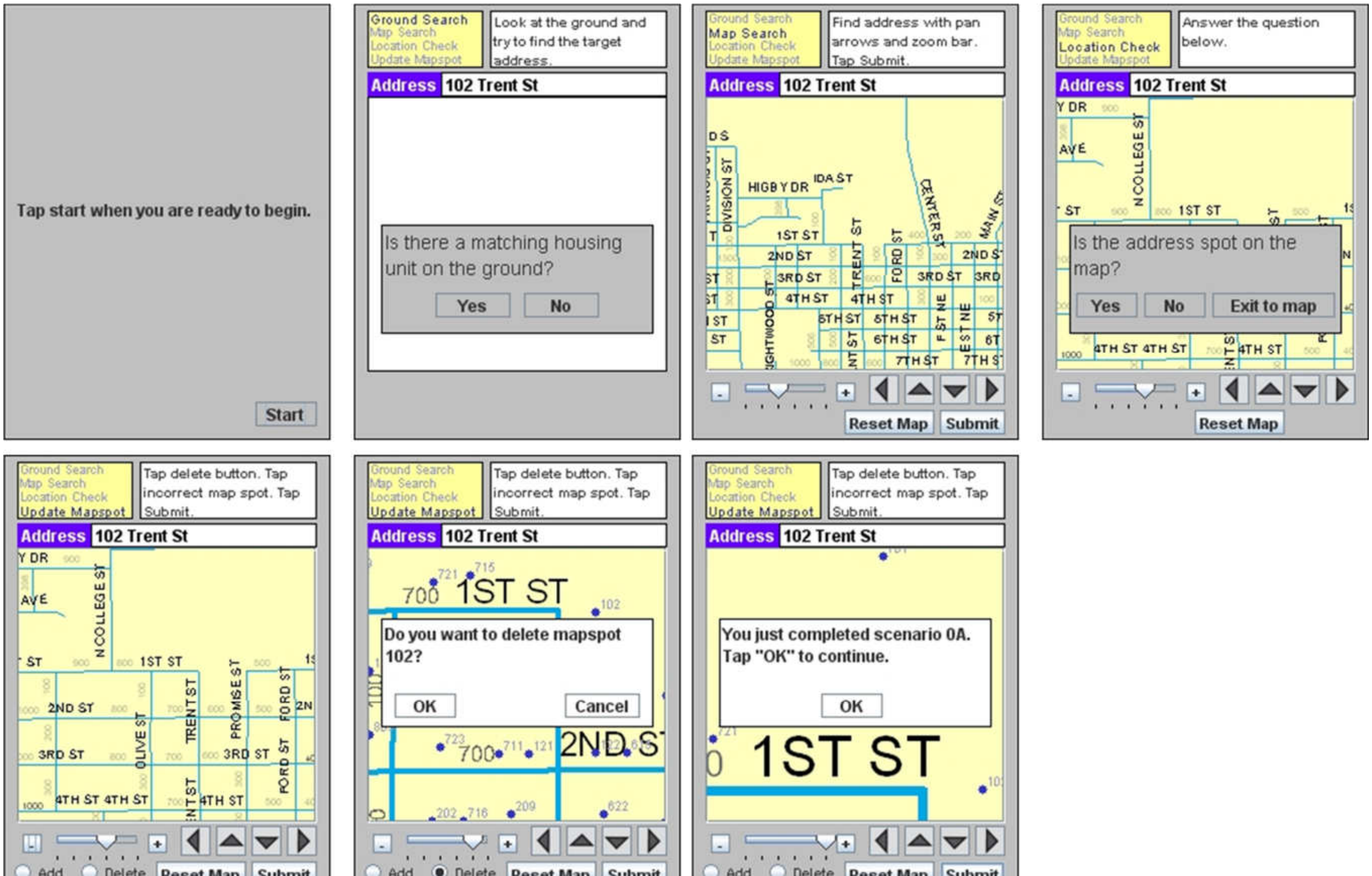


Figure 3.4 Guided interface (left) and unguided interface (right) (figure taken from Rusch, Nusser, Miller, Batinov, & Whiney (2012).

Figure 3.5 presents a sequence of screenshots of the address verification software with the guiding elements visible. The sequence depicts changes in the software interface as a participant proceeded through a type (b) scenario, “address needed to be deleted from the map.” The interface combined a map area (in the center) with a text display of the current target, pan and zoom buttons, “ add mapspot” and “delete mapspot” radio buttons, an “undo” button with the text “Reset Map”, a “Submit” button, and optionally, a step-by-step instruction list at the top of the interface. The software logged and timestamped all user interface actions (e.g. button clicks) as well as all mouse movement and events.

4



Tap start when you are ready to begin.
Start
Ground Search
Map Search
Location Check
Update Mapspot
Look at the ground and try to find the target address.
Address 102 Trent St
Is there a matching housing unit on the ground?
Yes
No
Find address with pan arrows and zoom bar. Tap Submit.
Reset Map
Submit
Answer the question below.
Is the address spot on the map?
Exit to map
Tap delete button. Tap incorrect map spot. Tap Submit.
Do you want to delete mapspot 102?
OK
Cancel
You just completed scenario 0A. Tap "OK" to continue.
Add
Delete

Figure 3.5 Address verification workflow: screenshots proceed from left to right and then down.

The real-world photos in the experiment were manipulated to reflect actual scenarios that address verification employees would encounter in the field. For example, a building was removed from a photo that was present on the software map, to simulate a situation where the building had been razed between consecutive Census surveys. Another manipulation was inserting label overlays on photographed streets and buildings to make the names and addresses of objects in the photo obvious to the participant. The scenarios varied along six factors: photo, street name, road configuration (four-way intersection, three-way intersection, and others), map location, and user facing depicted on the photo (north, south, east, west).

The map used in the experiment was assembled in ESRI ArcGIS from data layers of (1) streets, provided by the Census Bureau TIGER/Line servers at http://www.census.gov/geo/maps-data/data/tiger-line.html, and (2) address locations, provided by the Iowa Department of Transportation. The resulting map, visible in figures 3.4 and 3.5, approximates the visual presentation used by Census Bureau survey takers in the field. The Census Bureau map was, naturally, not available to us for the experiment, due to privacy concerns.

3.2.3 Results

Ordinary-least-squares regression identified patterns of interdependence between participant performance and cognitive test scores. There was statistical significance in relationships involving both spatial visualization (VZ) and perspective-taking (PT) abilities. To gain more understanding of the relationship between spatial visualization and perspective taking, two extra variables were constructed to capture different aspects of the co-variability of the two predictors. While perspective-taking ability was not interesting on its own, it became relevant in combinations with spatial visualization ability (Table 3.3). For that reason, the report contains three sets of analyses: (a) one with spatial visualization only; (b) one with spatial visualization and the difference VZ-PT, which captures the "effect" of having lower perspective-taking; and (c) an analysis with the average of (VZ+PT)/2 together with the difference VZ-PT, where the first term captures synergistic patterns of the two psychometric scores, and the second term captures the "effect of the gap".

Table 3.3 P values for ANOVA F-tests for performance and behavioral variables.

Variable	Time (s)	Accuracy (log m)
Spatial Visualization (VZ)	0.001[b]	
Spatial Difference (VZ-PT)	0.006[b]	0.02[c]
Spatial Average (VZ+PT)/2	0.05[c]	

a) Model only with VZ.
b) Model with both VZ and Spatial Difference.
c) Model with both Spatial Average and Spatial Difference.

The following response variables were tested: total time in seconds and accuracy of target address placement in m (log-transformed and thus allowing percentage interpretation). With both time and accuracy, lower scores (in seconds and log meters) indicated better performance.

The significant predictor variables included:

- ❖ Age, as a factor variable with levels 18-29, 30-39, 40-49, 50-59, and 60 and over;
- ❖ Gender as a factor variable (0 = male, 1 = female);
- ❖ Gender*Interface, as an interaction factor variable capturing difference in male-female response to guided vs. unguided interfaces;
- ❖ Spatial Visualization (VZ) as a numeric predictor;
- ❖ Spatial Difference (VZ-PT) as a numeric predictor measuring "the gap"; and
- ❖ Spatial Average ((VZ+PT)/2) as a numeric predictor measuring synergy.

In the analysis including both the spatial average and spatial difference, the average of visualization and perspective taking was statistically linked to total time. The coefficient estimate was -320 (SE=145). The coefficient is interpreted as follows: for every two points in either spatial visualization or perspective taking ability, the participant spent 320 seconds less on the exercise! The analysis with spatial visualization and spatial difference had a significant coefficient for spatial visualization of -327 (SE=189) and a significant coefficient for spatial difference of -245 (SE=74), indicating similar considerable increases in time performance.

The analysis including both spatial average and spatial difference revealed a negative relationship between spatial difference and housing unit location (in log meters). The value of the coefficient was -.69 (SE=.26), which is interpreted as follows: for every point of difference between spatial visualization and perspective taking, the user-determined housing unit locations was 69% farther from target location.

3.2.4 Conclusion

The Stationary Simulation study unearthed evidence of performance differentials on a computerized address verification task with a constrained workflow. The outcomes from the Stationary Simulation experiment supplement the outcomes from the "baseline" Paper Map study. The existence of statistically significant performance measures encouraged us to search for behavioral differentials. Behavioral differentials became the backbone of the detection technique.

3.3 Field and Virtual Reality Study

In the third address verification experiment, participants verified addresses with a handheld device in both the field and a high-fidelity immersive virtual environment. This new design built on both the freeform nature of the paper map study and the experience with software-aided workflows acquired in the Stationary Simulation study. The key features of the design were (a) participants' ability to freely move inside the experimental area, and (b) an interface which allowed completing scenarios in any order and resubmitting answers at will. Like previous studies, the experiment contained a cognitive testing phase and an exercise phase.

3.3.1 Team roles

This research study was conducted in collaboration with Kofi Whitney, Les Miller, and Sarah Nusser. Drs. Miller and Nusser acted as faculty advisors, while Kofi Whitney and Georgi Batinov shared the work in designing and executing the study equally.

3.3.2 Design

3.3.2.1 Recruitment, Compensation, and Compliance

Participants were recruited through flyers posted on the Iowa State University campus, and public bulletin boards in grocery stores and churches in Ames, IA. Participants were also recruited through a posting on the computerized online Student job board maintained by Iowa State University. The compensation offered was a $10 Target Gift card for participating in the cognitive testing phase, and $20 for participating in the field phase. Completion of the phases was not necessary for compensation to be offered. Participants were apprised of their rights in the experiment through a standardized Informed Consent form (see APPENDIX M).

3.3.2.2 Cognitive Testing Phase

During the cognitive testing phase, one-hundred-and-twenty-four participants were individually assessed on spatial visualization, visual memory, and perspective-taking ability. The tests were VZ-2, MV-2, and P-2 by Esktrom et al. (1976), and the perspective-taking assessment in Kozhevnikov et al. (2006). The location of the experiment was in an office on the Iowa State University campus. The cognitive testing phase lasted approximately an hour. At the beginning of the phase, participants had to read and sign an informed consent form. Tests were administered in immediate succession, with one-minute breaks between test sections and three-to-five-minute breaks between tests.

Participants with spatial visualization scores greater than or equal to 15 or less than 9 (out of 20) were randomly assigned to one of two treatments in the exercise phase. Pairs from either the low or high spatial visualization groups were randomized together, allowing each participant a .5 probability of assignment to either the virtual reality treatment or the field treatment. Thirty-two participants (14 males) were admitted to the second phase of the experiment.

3.3.2.3 Exercise phase – Field Treatment

For the field treatment, 15 participants (8 males) were taken individually to the exact same spot in a residential neighborhood in Ames, Iowa. They were first trained on using the handheld device, locating addresses in the field, and the think-aloud protocol. An observer provided them with a stylus and a handheld computer: a Pharos Traveler 535x with a 240x320, 3.5” transflective screen and a 624 MHz Intel PXA270 processor. The observer explained the task: determining whether a list of six addresses was correctly reflected on a software map (shown in Figure 3.6). Participants would have to physically walk to an address in order to answer the question. If the map contained errors, they had to use the software’s editing features to position the address at the correct location or remove it altogether. Four outcomes were possible. An address needed to either be added to the map, deleted, moved to a new location, or confirmed without changing the map. Participants were told to only correct the requested addresses and to ignore other possible errors on the map. The map contained no errors outside of scenario addresses (Figure 3.7).

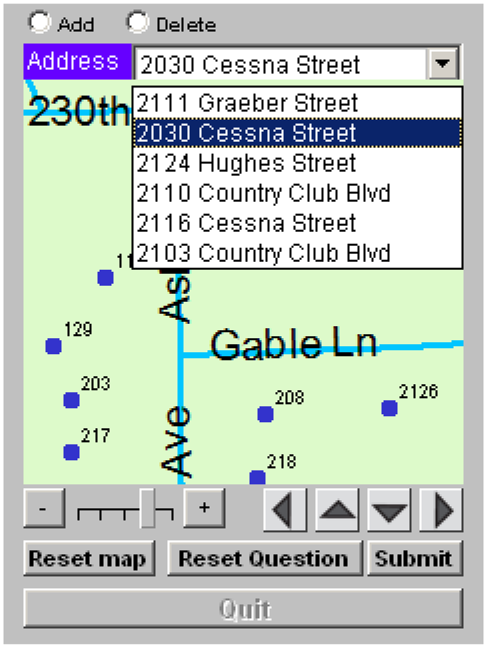


Figure 3.6 Address verification software with address list extended

Participants were then taught how to navigate and edit the software map, and were also instructed to verbalize all their thoughts for a think-aloud protocol. The map software was started in training mode and participants were asked to locate and verify three training addresses in the immediate vicinity, while the observers answered procedure questions and provided feedback on the quality of the think-aloud. At the end of the training session, observers answered the participant's final questions, and also explained that observers would not talk during the actual exercise or answer questions, other than to prompt the participant to keep verbalizing or to ask about behavioral details.

Observers then returned the participant to the exact location where all trainees started, switched the map software to experiment mode, and started an audio recorder (worn by the participant) and a GPS tracker (carried by the observer). The GPS tracker was not given to the participant so that they would not be interrupted to time-stamp scenario completions. In return, observers shadowed the participant, establishing a close approximation of the exercise path. Figure 3.7 depicts an example participant in the field.

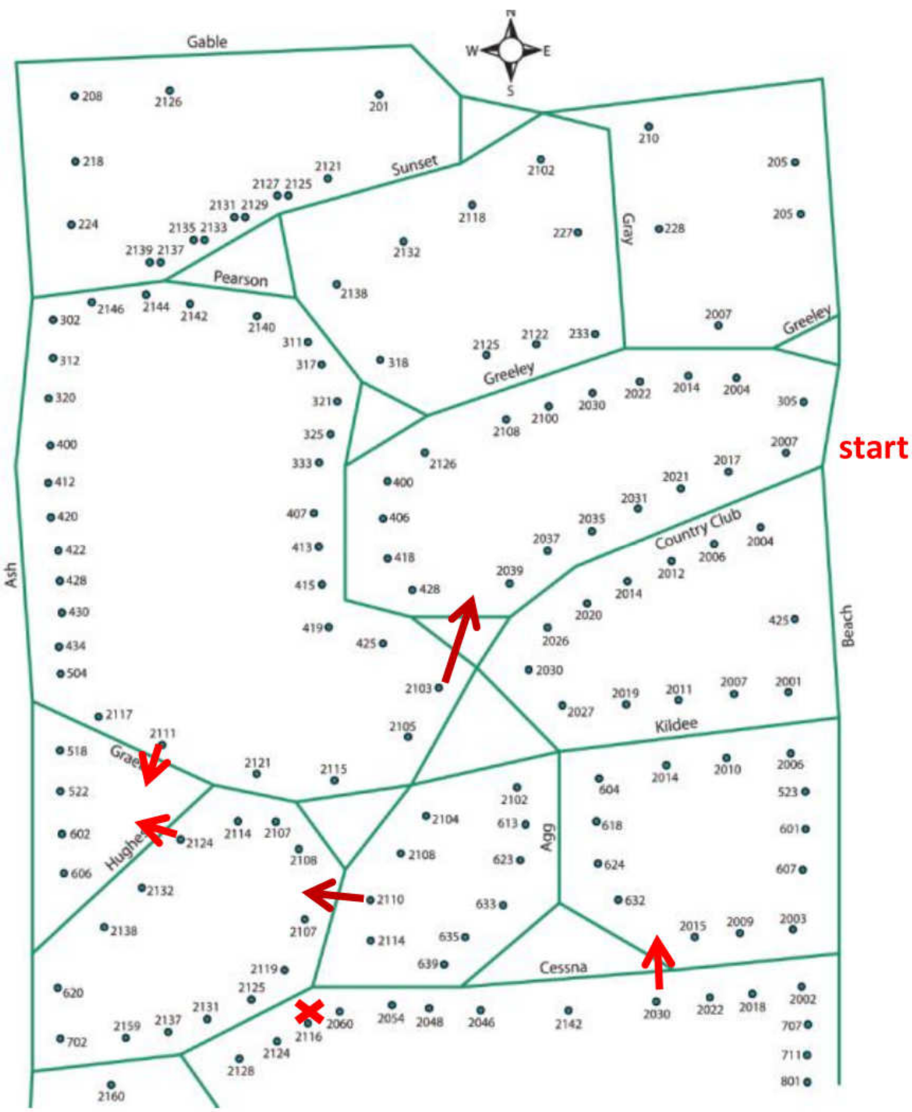


Figure 3.7 A correct map together with the errors introduced to six target addresses. 2111 Graeber St, 2124 Hughes Ave, 2110 Country Club Blvd, 2103 Country Club Blvd, and 2030 Cessna St were moved to an incorrect location, while 2116 Country Club Blvd was deleted.

All exercise-takers verified the same six addresses from an identical randomized list order (see Figure 3.7 for a complete list of targets together with their error status), and therefore could not benefit from inadvertent route hints. The list could be invoked at all times in software by tapping the currently selected scenario (Figure 3.6). Participants were allowed to work on addresses in any order and could return to previously submitted scenarios as many times as they wanted. Only final answers were evaluated for correctness.

After participants solved their final scenario, they could signal that they had finished the exercise.

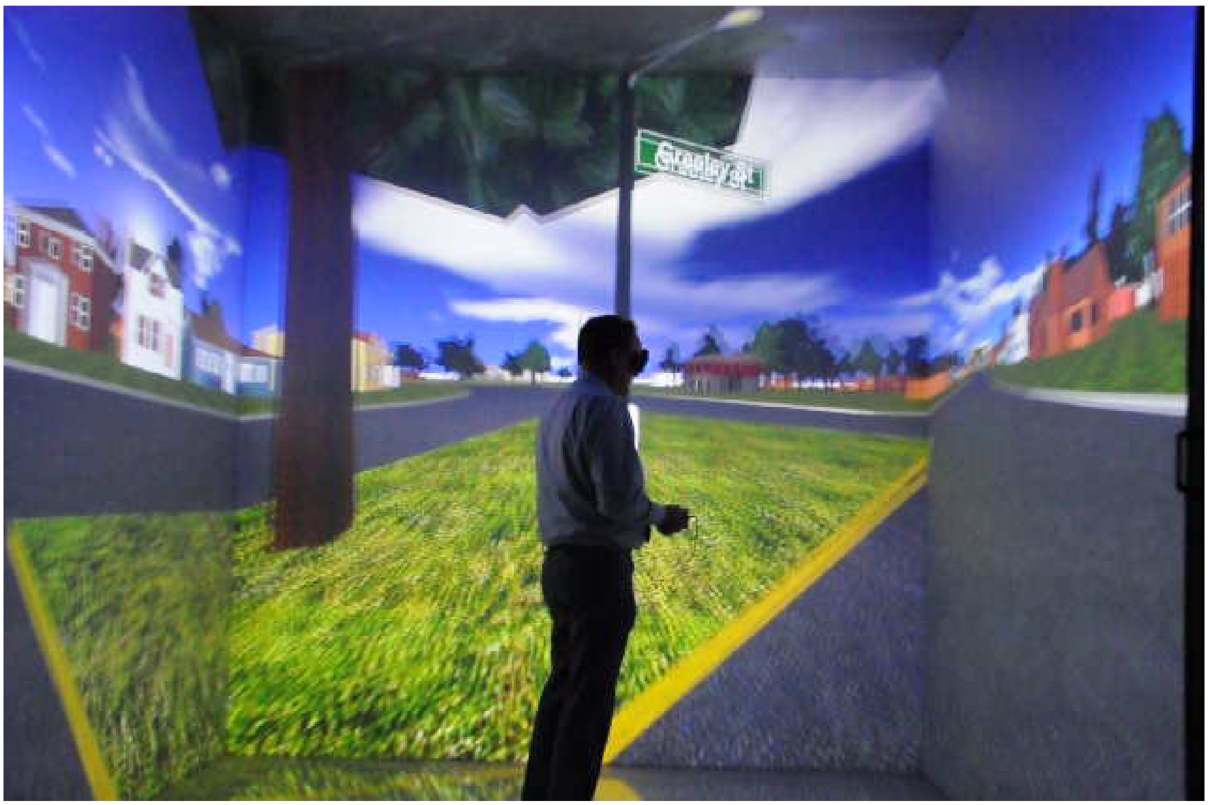


Figure 3.8 Dr. Les Miller inside the C6 immersive virtual environment. Five of the six projection walls of the environment are visible. The sixth wall is retracted to expose the participant for the shoot. A street sign of Greeley Street is in the foreground. The stereoscopic double image allows for depth perception when the user is wearing stereo glasses (pictured).

<u>3.3.2.4 Exercise phase – Virtual Reality Treatment</u>

Seventeen participants (6 males) were randomly assigned to the virtual reality treatment and were taken individually to a C6 immersive virtual reality environment on the Iowa State University campus (Figure 3.8).

3.3.2.4.1 Virtual reality model

The virtual setting loaded in the environment was a high-fidelity three-dimensional model of the residential area (Figure 3.9), with one more block modeled outside the westernmost and easternmost extents of the map. The dimensions of the model were roughly 600 x 600 m (2000 x 2000 feet). The model was created in SketchUp (http://www.sketchup.com) and imported into the virtual reality environment through VR Juggler (http://www.vrjuggler.org). Housing units and streets were georeferenced. Actual housing units were represented by house models of similar size and style selected from Sketchup's repository of three-dimensional housing models (http://sketchup.google.com/3dwarehouse/). The neighborhood model also incorporated notable landmarks in the area, street signs, curbs, textured surfaces, a day sky with sun, and trees and shrubs. Multi-lane streets and split boulevards were represented correctly. The model did not include sidewalks.

3.3.2.4.2 Virtual reality equipment

The virtual reality room was a cube with dimensions 3.05 x 3.05 x 3.05 m. Each of the six walls displayed stereo images of 4096 x 4096 pixels at approximately 16 frames per second. Video projection was driven by a cluster of 48 HP xw9300 workstations with 96 nVidia Quadro graphics cards sending video frames to 24 Sony SRX-S105 digital cinema projectors. InterSense's IS-900 tracking system tracked the participant's head location and gaze direction, and the stereo perspective dynamically shifted with the user's gaze. The participant wore active stereo glasses.

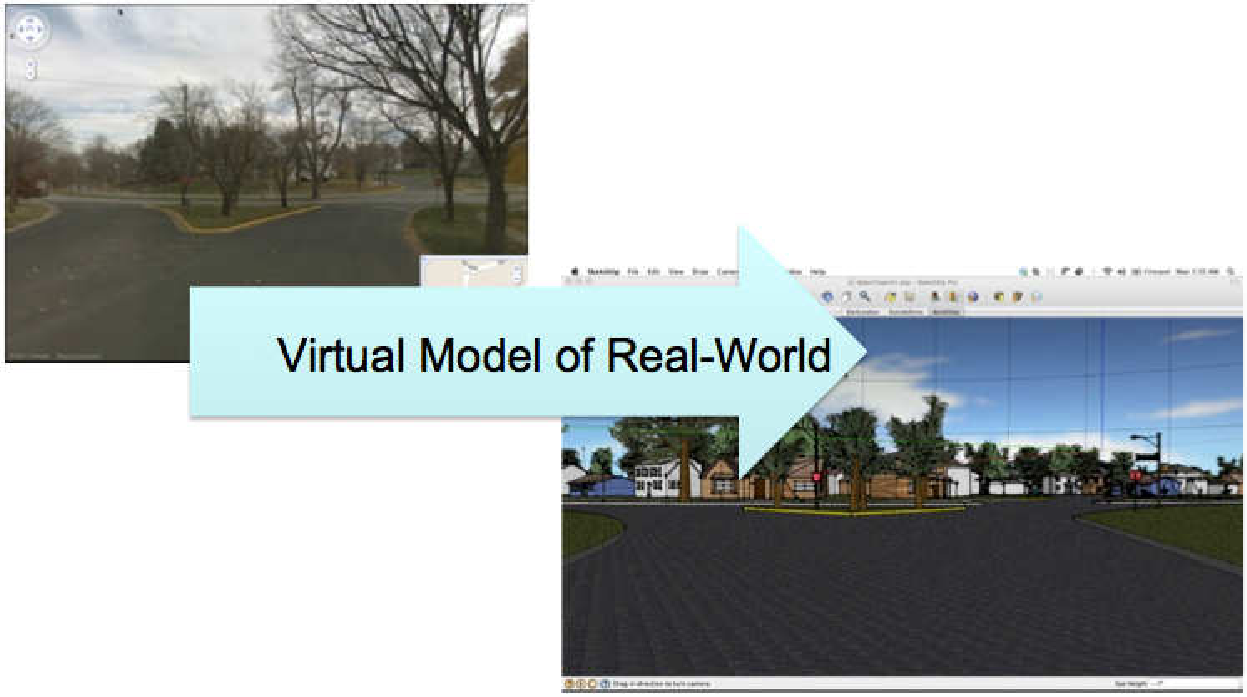


Figure 3.9 The virtual environment model was a high-fidelity replica of an Ames neighborhood.

3.3.2.4.3 Moving in virtual reality

Movement in the environment was accomplished by stepping towards the desired direction. A circular spot in the center of the floor, approximately 0.6 m (24 in) in diameter, was the “dead zone”. If the participant’s head was located in the column of the spot, all movement stopped. Stepping outside the dead zone would start moving the virtual reality model in the opposite direction of the step, giving the illusion of the participant moving through the model in the direction of the step. As the participant stepped closer to the walls, movement speed increased, from approximately 0.1 m/s to a maximum of approximately 2.22 m/s (8 km/h or 5 mi/h). The maximum speed was set to a slow trot because of concerns that a higher speed could not be encountered in the range of walking speeds available to participants in the field treatment, and a lower maximum speed could bore participants, causing them to lose focus.

3.3.2.4.4 Protocol differences from the Field treatment

The exact same protocol was employed for both treatments, with one exception. Prior to introducing the handheld device, participants were trained on moving inside the virtual environment.

3.3.3 Results

The results in this section were published in Batinov, Whitney, Miller, Nusser, Stanfill, & Ashenfelter (2013). When compared to the Paper Map Study, The Field-VR study had a similar, albeit more complicated, design. Taking into account the findings from the Paper Map and Stationary Simulation studies, we postulated the following two performance hypotheses:

Hypothesis 1:	High-spatial-visualization participants would travel significantly shorter distances than low-spatial-visualization participants.
Hypothesis 2:	High-spatial-visualization participants would take less time than low-spatial-visualization participants in both the field and virtual environments.

To accommodate the increased complexity of the experiment, the statistical tool of choice was ordinary-least-squares regression. The model took the following form:

(1) $Y = E + S + E*S + G,$

where:

- Y is the response variable (a performance metric, log(distance) and log(time));
- E is a factor variable denoting environment type (0 = field, 1 = virtual);
- S is a factor variable denoting spatial visualization ability (0 = high; 1 = low);
- E*S is an interaction of the environment and spatial levels; and
- G is a factor variable for gender (0 = female; 1 = male).

The log form of the response variable allows for easy interpretation of regression coefficient: a coefficient, when multiplied by 100, describes the percentage change in the response variable that is attributable to a change of level of the predictor variable. Illustrative examples from the actual data will be presented in short order.

The *Environment*Spatial* interaction term is attempting to capture differences in performance of a given spatial ability level when the environment varies. What that term adds to the model is accounting for possibility of high-spatial participants having greater or smaller performance differential from low-spatial participants when in virtual reality, as compared to the differential in the field.

Table 3.4 presents the results where time is the response variable.

Table 3.4 Regression results for log(time).

Term	Estimate	Std. Error	t-value	Pr(>|t|)
(intercept)	3.338	0.112	29.814	0.00
Environment	0.083	0.159	0.524	0.605
Spatial visualization	0.398	0.152	2.622	0.014
Gender	0.005	0.107	0.050	0.961
Env*Spatial	0.014	0.209	0.068	0.946

Of the tested variables, only spatial visualization proved to be a significant predictor of time performance. The log form of the response allows us to state the significant result in the following form: low spatial visualization participants, on average, took 39.8% more time to complete the exercise. The other potential predictors showed no evidence of gender or environment affecting the time performance of participants.

Table 3.5 shows outcomes for the ordinary-least-squares regression model where distance is the response variable.

Table 3.5 Regression results for log(distance).

Term	Estimate	Std. Error	t-value	Pr(>|t|)
(intercept)	0.021	0.117	0.177	0.861
Environment	0.268	0.166	1.617	0.117
Spatial visualization	0.373	0.159	2.355	0.026
Gender	0.029	0.111	0.261	0.796
Env*Spatial	-0.425	0.218	-1.944	0.062

The log form of the response variable allows us once again to state the performance result as a percentage. The model suggests that low-spatial participants, on average, travelled 37.3% longer distances while completing the exercise. The other predictors – environment and gender – once again failed to reach significance levels. Tables 3.4 and 3.5 show that both performance hypotheses were validated.

3.3.4 Conclusion

The virtual reality and field treatments presented a new set of environments, new locations, new software, and a new protocol to extend our understanding of the relationship between spatial ability and computer behaviors. In this last and most elaborate of the three experiments, performance differentials were once again linked to spatial visualization ability. The existence of performance statistics encouraged the search for behavioral statistics, which would then promote behaviors for spatial visualization detection.

3.4 Conclusion

Behavioral differentials are crucial to the proposed dissertation as the backbone of any approach to automatic recognition of spatial visualization. This chapter reported on our investigations of address verification in three distinct experiments: the “baseline” Paper Map study, the “constrained workflow” Stationary Simulation study, and the “unconstrained workflow” Field and VR study. All three research attempts produced statistical evidence for the divergence in performance between high-spatial-visualization and low-spatial-visualization participants. These quantitative outcomes were indicators of the potential of behavioral differentials, which empowered automated ability recognition. As a logical next step, the coming chapter reports on behavioral statistics.

CHAPTER 4. BEHAVIORAL STATISTICS

Chapter 3 presented statistically significant performance differentials between high-spatial-visualization and low-spatial-visualization participants in all three address verification experiments. The availability of statistical results in performance encouraged a search for systematic behaviors that can inform computerized spatial visualization detection. This chapter reports on statistical evidence of behaviors in the three studies.

4.1 Paper Map Study

The Paper Map Study was described in detail in Chapter 3. Data for the statistical tests on behavior came from three sources: user notes, observer notes, and think-aloud protocols. Data from user notes consisted of the number and classifications of the marks made by users on the provided paper map and address list. Observer notes contained the number and descriptions of behaviors exhibited by users. Think-aloud protocols were audio recordings of participants who verbalized their thoughts as they confronted the exercise. The recordings were encoded into a rich set of events that could support statistical queries. The set of unique think-aloud codes can be found in APPENDIX H.

4.1.1 Annotation behaviors

The correlations of cognitive test scores and annotating behaviors of participants were tested. Four behaviors exhibited significant correlations to cognitive ability (Table 4.1).

Table 4.1 Association of cognitive test scores with map and list variables (Welch's *t* test)

Variable	**Cognitive Test**	$\overline{Y}_1$- $\overline{Y}_0$*	$SE(\overline{Y}_1$- $\overline{Y}_0)$**	***p***
Target streets highlighted on map	Spatial Visualization	-4.25	1.54	0.01
Map annotations	Perspective-taking	-4.45	1.52	0.02
List annotations	Perspective-taking	-4.31	1.90	0.05
Route sequence on list	Visual Memory	3.29	1.11	0.01

* $\overline{Y}_1$ is the mean of cognitive test scores for all who exhibited the behavior.
**$\overline{Y}_0$ is the mean of cognitive test scores for all who did not exhibit the behavior.

Participants with lower spatial visualization scores tended to highlight target streets on the map. Additionally, participants with lower perspective-taking scores tended to leave (1) more marks on the map and, (2) more marks on the list of addresses. The above three behavioral differences provide evidence that lower-spatial-ability participants desired and created more visual workflow elements. This outcome resonates with the findings of Jennings, Benyon and Murray (1991), where high-spatial-ability participants performed better with a command-line interface.

In the fourth behavioral difference, participants with higher visual memory were more likely to record an ordering of visited addresses. This result superficially appears counter-intuitive because of the previous three behaviors. However, the sequences were written on the list of addresses on the back of the clipboard, which prevented users from viewing the map concurrently with the list. All but one user (who detached the back sheet) were forced to flip back and forth between map and list. High-visual-memory participants would have an advantage at recording the route sequence without looking at the map.

4.1.2 Observer-reported behaviors

More statistically significant behavioral differences were present in user behaviors reported by observers. The data for this statistical test consisted of the number of observed occurrences of a particular behavior, as recorded by study administrators. Spatial visualization test scores were positively correlated with *address error pre-detection* ($n = 21$, $r = 0.44$, $p = 0.05$), which tracked user tendency to discover map errors during the map inspection at the beginning of the exercise. This correlation suggests high-spatial users made inferences about target correctness based on reading map detail alone.

Spatial visualization scores were also positively correlated with *nearest address selection* ($n = 21$, $r = 0.45$, $p = 0.04$), which tracked user tendency to choose the closest available address when selecting verification targets. This correlation suggests high-spatial users minimized distance traveled in the short term: a strategy which appears to have contributed to the improved performance of the high-spatial group.

Perspective-taking test scores were positively correlated with *address error pre-detection* ($n = 25$, $r = 0.49$, $p = 0.01$) and *cardinal heading usage* ($n = 23$, $r = 0.51$, $p = 0.01$), indicating that participants with higher perspective-taking scores found target address errors during initial map inspection, and also verbalized a cardinal (north-south-east-west) frame of reference.

4.1.3 Phase-specific behaviors

4.1.3.1 Workflow phases

Two broad workflow phases were distinguishable in the verification of a single address: the "approach" phase and the "verification" phase. The user was in the "verification" phase when in a physical vicinity of the target housing unit that contained enough information to verify the address. The user was in the "approach" phase while navigating to the target vicinity.

The approach and verification phases were distinguished as follows. The vicinity of a target housing unit included two immediate neighbors on the left, two immediate neighbors on the right, and three immediate neighbors on the opposing side of the street. The approach phase ended when a participant verified a neighboring address on the ground.

4.1.3.2 Think-aloud protocols

Think-aloud protocols were audio recordings of participants who verbalized their thoughts as they confronted the exercise. The recordings were encoded into a rich set of events that could support statistical queries. The set of unique think-aloud codes can be found in APPENDIX H.

Analysis of the protocols yielded a number of significant behavioral differences between participants of low spatial ability and high spatial ability. The findings from the approach and verification phases are discussed in their own sections.

4.1.3.3 Approach phase

Eleven statistically significant behavioral differences were discovered through Wilcoxon-Mann-Whitney two-tailed tests during the approach stage (Table II). Four of these behaviors were only exhibited by the low-spatial group.

TABLE 4.2 BEHAVIORAL DIFFERENCES - APPROACH STAGE (WILCOXON-MANN-WHITNEY TWO-TAILED TEST)

Behavior	**Group that is more likely to exhibit the behavior**	**p-value**	**Only one group exhibits the behavior**
Heading selected	Low	0.00	Yes
Identified map relation erroneously	Low	0.045	Yes
Numbering pattern recognized	Low	0.02	
Map relation identified	Low	0.04	
Map rotated	High	0.00	
Navigation plan reinforced	Low	0.02	
Planning with a map	Low	0.01	Yes
Orient self with regard to cardinal directions	Low	0.00	Yes
Position located on map	Low	0.02	
Recall target	Low	0.0497	
Street identified	Low	0.02	

The only behavior that was exhibited more frequently by the high-ability group was map rotation (“map rotated”, n = 21, p = 0.00), in which the participant would turn the map in a two-dimensional plane roughly perpendicular to their gaze incidence, in order to align it with either their facing direction (and thus obtain a “track-up” view of the map) or with known elements in reality, for example, when determining if an address was in the correct location.

Low-spatial participants were more likely to verbalize planning with a map (“planning with a map”, n = 21, p = 0.01). We found no statistical evidence for difference in verbosity between low-spatial and high-spatial participants. This means low-spatial participants spent more time planning with the map.

Continuing the pattern of map behaviors, low-spatial participants were more likely to speak out relations inferred from the map (“map relation identified”, n = 21, p = 0.04). At the same time, low-spatial users were the only group to make observable mistakes while decoding the map (“identified map relation erroneously”, n = 21, p = 0.045). This behavioral differential cannot be attributed to spending more time with the map, because map usage was also tracked with the very frequent “check map” event, which was not statistically significant between the two groups.

Low-spatial participants were also more likely to verbalize recognizing street numbering patterns, which included odd-or-even sides of the street and directions of number increase or decrease (“numbering pattern recognized”, n = 21, p = 0.02). The low-spatial group exhibited a behavior called “orient self with regard to cardinal directions”, whereby a participant would convey having aligned themselves along a north-south-east-west frame of reference (n = 21, p = 0.00). An example of the behavior would be the statement, “Facing west, Ash Avenue is in front of me.” This finding needs to be contrasted with the cardinal-usage differential reported in the “Observer-Reported Behaviors” section (n = 23, p = 0.01), where high-perspective-taking participants were more likely to use a cardinal frame of reference. The two findings are not contradictory, as explicit self-alignment was only present in the low-spatial group, while cardinal direction usage was encountered in both groups.

The “heading selected” event denoted a participant declaring an immediate direction of movement using egocentric or geographic frame of reference such as “I am going to turn left” or “I am heading south”. Low-spatial participants selected a heading more frequently ($n = 21$, $p = 0.00$).

Low-spatial participants were also more likely to verbalize refinements to their navigation plans, as evidenced by the “navigation plan reinforced” variable ($n = 21$, $p = 0.02$). For a navigation pronouncement to be considered a plan, it needed to contain at least two segments, such as, “I will take Ash and then Pearson to get to Greeley”, or a segment and two turns, such as “I will turn right on Ash and then turn left after two blocks to get to the address”. Shorter navigation pronouncements were not considered elaborate enough to constitute planning. Yet statistics in Chapter 3 showed that low-spatial participants took a longer route to complete the exercise, and we already saw that low-spatial users tended to select a heading more frequently. Both of these behaviors show that the low-spatial group exhibited lower planning efficiency.

Low-spatial participants mentioned street names more often (“street identified”, $p = 0.02$), recalled their previously chosen target address or street more often (“recall target”, $p = 0.0497$), and located their own position on the map more often (“position located on map”, $p = 0.02$). All three behavioral differentials point to less efficient interaction with the map. Street names were the second most common cue category after address numbers. However, there was a limited set of streets within the exercise area, so low-spatial participants repeated street names more often. Unlike address numbers, which did not have complex relationships to one another, street relations formed the strategic layout of the exercise. Further, the experiment design used a complex, non-uniform street layout. The inefficiency in assembling cues, as portrayed by the "street identified" variable, was accompanied by ongoing efforts to keep refreshing the mental model, as evidenced by low-spatial users more frequently recalling their previously chosen targets and more frequently determining their own location.

4.1.3.4 Verification phase

During the verification stage, the high-spatial-visualization group was more likely to rotate the map (Wilcoxon-Mann-Whitney two-tailed test, $n = 21$, $p = 0.01$). When rotating the map, a participant would change the orientation of the map in a two-dimensional plane roughly perpendicular to their gaze incidence, so they could look at the spatial configuration from a different angle. Participants would sometimes state that they rotated the map to align it with the direction they were currently facing, thereby using a "track-up" map view. Another reason they rotated the map was to match the direction of the target address configuration on the map with reality in order to solve the scenario.

4.1.3.5 Summary of phase-specific behaviors

The phase-specific behavior outcomes were, as a whole, non-intuitive. A naïve expectation of group behavior differentials would anticipate the high-ability participants to exhibit more strategies, on average, while engaging the exercise, but a more complex picture of behavioral differences emerged: low-spatial participants engaged a set of strategies more often than their high-ability counterparts, with higher reported incidence, but lower effectiveness. Figure 4.1 represents the differential decision models for low-spatial and high-spatial users in the Paper Map study.

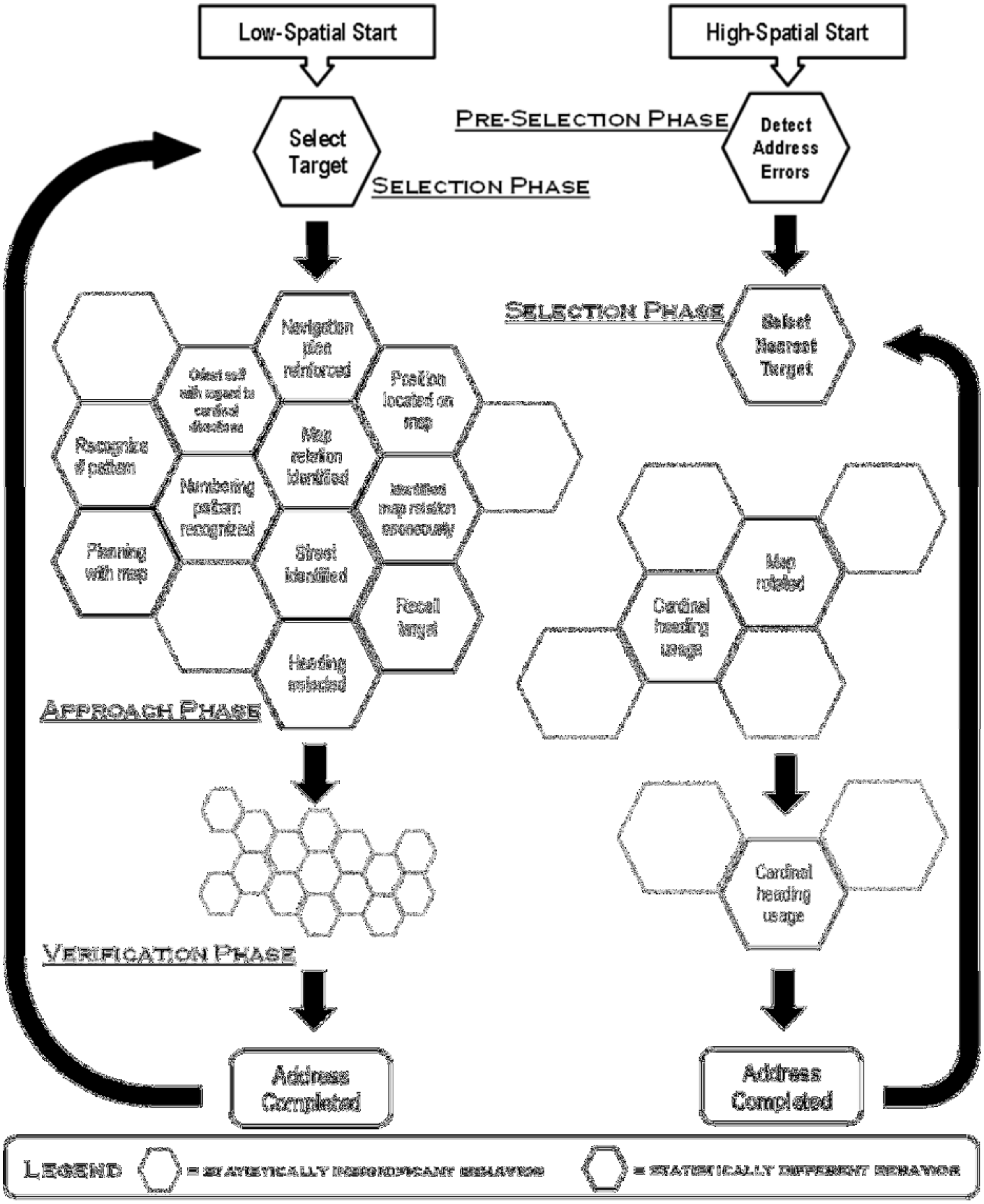


Figure 4.1 Decision Models for the Paper Map study.

4.1.4 Conclusion

Behavioral analysis discovered systematic differences between the low-spatial and high-spatial groups. The “baseline” Paper Map experiment encouraged further research into behaviors for automated ability detection.

4.2 Stationary Simulation Study

The behavioral results in this section were previously published in Rusch, Nusser, Miller, Batinov, & Whitney (2012). The study design was described in the previous chapter.

Ordinary-least-squares regression was used to account for the variation in number of zoom actions, number of software map resets, and number of pan actions. All three software behaviors were significant (Table III).

Table 4.3 Stationary Simulation Study: ordinary-least-squares regression p-values for behavior.

Variable	Number of zoom actions*	Number of map reset actions**	Number of pan actions***
Spatial Visualization (VZ)	0.03[b]	0.03[a]	0.03[a]
Spatial Difference (VZ-PT)			0.02[c]
Spatial Average (VZ+PT)/2			

a) Model only with VZ.
b) Model with both VZ and Spatial Difference.
c) Model with both Spatial Average and Spatial Difference.
* Number of times the map view was switched between lower- and higher-scale versions.
** Number of times the map view was returned to its initial geographic coordinates and scale.
*** Number of times the user clicked a button to move the map view a set distance to the north, south, east, or west.

The number of zoom actions is defined as the number of times the user switched the map view between lower- and higher-scale versions. Clicking the “zoom-in” button, labeled in the software with a “+” sign) switched the view to a lower-scale (higher-detail) version of the map while remaining centered on the currently observed area. Clicking on the “zoom-out” button (labeled in the software with a “-“ sign) changed the view to a higher-scale (lower-detail) version of the map while remaining centered on the currently observed area.

In the analysis with spatial visualization ability and spatial difference, spatial visualization ability was negatively associated with zoom, at a coefficient of -5.56 and a standard error of 2.36, suggesting that each extra point in spatial ability was associated with five and a half fewer zoom actions.

Map reset actions occurred when participants clicked the “Reset map” button. The map view was returned to its initial geographic coordinates and initial scale. These coordinates were different for each of the ten experimental scenarios. In the analysis with spatial visualization only, spatial visualization ability was found to be negatively associated with number of map resets, with a coefficient of -1.93 and a standard error of 0.82. The interpretation of this result is that for each extra point of spatial visualization ability, participants reset the map on 2 fewer occasions.

Pan actions were defined as the user clicking one of the “up”, “down”, “right”, and “left” buttons to move the map view by a set distance to the north, south, east, and west, while keeping the map scale constant. In the combined analysis with spatial visualization and spatial difference, number of pan actions was negatively associated with the difference between spatial visualization and perspective taking abilities (coefficient value = -52.13, standard error = 19.94), while in analysis with spatial visualization only, pan actions were significantly associated with spatial visualization. The statistical findings reinforced a perspective that people with relatively higher spatial visualization ability tended to pan around the map considerably less. This outcome harmonized with Paper Map Study findings that low-spatial participants were less efficient with the paper map.

This section presented evidence in support of the feasibility of automatic detection. Pan actions, zoom actions, and reset actions were all software events that could be utilized by a computer device to make decisions on whether a user possessed low or high spatial visualization ability. We expected to see these behaviors in actual detection algorithms.

4.2.1 Conclusion

Behavioral analysis in the Stationary Simulation provided direct evidence of differentiation between spatial visualization levels at the user interface. The outcomes of this study encouraged further work on automatic detection.

4.3 Field and VR study

The final address verification experiment was designed to observe participants with a handheld device verifying street addresses in both the field and a high-fidelity immersive virtual environment. This design built on both the freeform nature of the paper map study and the experience with software-aided workflows acquired in the Stationary Simulation study, with new handheld hardware and software that was designed relaxed all possible workflow constraints present in the Stationary Simulation experiment. The key features of the third experiment were (a) participants' ability to free-roam as they found addresses to verify; (b) all work was performed on a small PDA-style handheld computer; (c) the graphical user interface allowed completing scenarios in any order and resubmitting answers at will; and (d), data was acquired from both the real world and a high-fidelity immersive virtual environment. Like the previous studies, the experiment contained a cognitive testing phase and an exercise phase. For complete details on the experimental setup, please refer to Chapter 3.

Ordinary-least-squares regression revealed significant behavioral coefficients. The results in this section were published in Batinov, Whitney, Miller, Nusser, Stanfill, & Ashenfelter (2013).

Table 4.4 Ordinary-least-squares regression results for log(time) as the response variable

Term	Estimate	Std. Error	t-value	Pr(>|t|)
(Intercept)	3.111	0.120	25.979	0.000**
Environment	0.041	0.138	0.297	0.769
Spatial	0.301	0.134	2.235	0.034**
Resetpans	0.036	0.011	3.213	0.003**

*Marginally significant at p < 0.10.
**Statistically significant at p < 0.05.
***Statistically significant at p<0.01.

In the regression of Table 4.4, both spatial ability and the Resetpans variable were significant. Resetpans was a numeric variable equal to 0.5(Resets-mean(Resets)/sd(Resets) + Pans-mean(Pans)/sd(Pans)). This was the mean of the normalized values of reset and pan actions performed by the user. Dividing by the standard deviation was used to normalize the contribution of resets and pans to the variable, because pans were more frequent than resets. The Resetpans coefficient indicated that participants took 3.6% more time to complete the exercise for each standard deviation of reset and pan actions.

A regression with log(Distance) is presented in Table 4.5.

Table 4.5 Ordinary-least-squares regression results for Log(Distance) as the response variable

Term	Estimate	Std. Error	t-value	Pr(>|t|)
(Intercept)	-0.200	0.128	-1.563	0.130
Environment	0.227	0.148	1.541	0.135
Spatial	0.278	0.144	1.932	0.064*
Resetpans	0.035	0.012	2.915	0.007***

*Statistically suggestive at p < 0.10.
**Statistically significant at p < 0.05.
***Statistically significant at p < 0.01.

The Resetpans variable was once again significant and positive, contributing an extra 3.5% to variability per standard deviation. The two potential explanations for this statistic are quite fascinating. Either participants were moving aimlessly as they were performing additional pan and reset actions, or they made wrong choices on traveling which they had to correct.

Overall, when we introduced software actions as predictors for user performance in the Field and Virtual Reality experiment, the combination of reset and pan actions was statistically significant. This outcome harmonized with the significance of software actions in the Stationary Simulation study and encouraged further progress.

4.4 Conclusion

In this chapter, we presented statistical analyses of user behavior from three experiments. We purposefully omitted any qualitative discussion of the observed strategies of participants, because qualitative observations do not constitute sufficient reason to instrument a software response. Quantitative findings, on the other hand, were critical to the detection effort.

The behavioral narrative in this chapter revealed that between-group differentials in spatial visualization persisted across the three experiments. Data transformations revealed statistically significant relationships involving user behaviors. The persistence of behavioral outcomes was the strongest indicator to continue research. The next chapter discusses the actual detection technique for the Stationary Simulation Study and the Field and Virtual Reality Study.

CHAPTER 5. AUTOMATIC DETECTION

Chapters 3 and 4 reported on statistical evidence for performance and behavioral differentiation between high- and low-spatial-visualization users on address verification tasks like those performed by the Bureau of Census. The observed statistical differentials suggested spatial visualization ability is detectable. The next step, therefore, was to attempt detection. This chapter will describe a successful approach. As discussed in Chapter 1 and Chapter 2, the goal was to achieve 80% or higher percent correct predictions. Such accuracy would improve on results in the current literature and be high enough to allow practical application.

An additional goal was to use no additional sensors, in order to facilitate adoption on generic devices in industry. Although frequently reported in the literature, sensor feedback, including galvanic skin response, eye tracking, or pressure detectors, was considered an impediment to adoption, for three reasons. First, sensors could require effort from users, as in the case of a galvanic skin response sensor, and therefore could be a "nuisance" to be avoided. Second, sensors are still in the process of becoming widespread and are not available on all devices. Third, the cost of additional hardware could impede large deployments.

The following sections present detection outcomes achieved on the Stationary Simulation Study and the Field and Virtual Reality Study. The results are encouraging: differences between the two experimental protocols did not prevent high detection rates, and therefore provided ecological validity to the detection approach. In particular, the Stationary Simulation Study had an especially restrictive protocol, while the Field and Virtual Reality study had an unconstrained protocol. In the Stationary Simulation Study, address verification was accompanied by an unavoidable sequence of questions, addresses had to be solved in a fixed order, participants could not move, and verification cues were pre-assembled on photographs. Conversely, in the

Field and Virtual Reality Study, there were no intermediate questions to answer, addresses could be verified repeatedly in any order, participants could navigate the environment at their leisure, and select their own verification cues. We achieved better than 80% accuracy in both studies, demonstrating the viability of the approach to dissimilar protocols and environments.

The chapter's first section discusses what features enabled successful detection: user map operations with or without geographical tags.[2] The section also provides a short example of the data representation, followed by the list of learning algorithms. Sections 5.2 and 5.3 present the detection outcomes for the Stationary Simulation and Field and Virtual Reality experiments, respectively. Section 5.4 examines the best-performing algorithms in detail to verify the learning models make sense. Section 5.5 outlines the contributions of the detection approach to current state of research, and Section 5.6 discusses expectations for future deployments. Section 5.7 concludes the chapter.

5.1 Feature Selection

5.1.1 The need for a low feature-to-instance ratio

Each participant log contained between 2,727 and 101,528 lines of text, thereby associating a myriad of features with each classification instance. A problem arose: when there are many features relative to classification instances in a corpus, machine learning algorithms struggle with interpreting the available information, a condition known as the "curse of dimensionality" (e.g. Blum and Langley 1997, pp. 245-246, Domingos 2012, p. 81-82). An intuitive explanation for this phenomenon is that each additional feature's marginal effect is to

[2] Chapter 6 shows that geographical distribution of behavior differentials in the Field and Virtual Reality Study mirrored the outcomes in the Paper Map Study.

explode the classification space that an instance needs to be mapped to. For example, from an algorithm's viewpoint, an instance with ten Boolean features, which are the simplest class of features, needs to be classified against up to 2^{10}, or more than a million, unknown instances. Contrastingly, an instance with twenty Boolean features needs to be classified against 2^{20}, or more than a billion, possible instances. The number of training observations (participants) shrinks relative to the feature space with each additional feature. A difficulty also arises from extra features that may be redundant or irrelevant. Additionally, in high dimensions, distributions do not resemble their low-dimensional counterparts and so impede both approximation and intuition: for example, a many-dimensional Gaussian distribution has almost all its weight in the tails (Blum and Langley 1997, pp. 245-246, Domingos 2012, p. 81-82).

To counteract the challenge of multiple features, we can search for pattern-rich subsets of features. These subsets are computationally and algorithmically easier to compare, but more importantly in our case, they were supposed to capture behavioral differences between participants. We already saw statistical analysis in previous chapters showing some map operations varied significantly with spatial visualization ability. Therefore, we mounted a detection effort based on tracking map operations.

There were four reasons why the set of map operations became the set of classification features. First, in our map-centered experimental software, they constituted the majority of interface affordances, and participants spent almost all their actions performing map operations. Second, zoom, pan and reset operations were statistically significant in the analysis of the Stationary Simulation and Field and Virtual Reality studies. Third, map operations had intuitive interpretations in terms of human behavior. Finally, protocol stages in the Stationary Simulation experiment could not be admitted as classification features, because they provided indirect knowledge about the state of the world. In particular, in the Stationary Simulation protocol, participants were presented with the following questions about the address: “Is the address on the ground?”, “Is the address on the map?”, “Is the address in the correct location?”, and if they answered incorrectly, they were forced to redo the stage preceding the question.

The exact forms of the transformations for both studies will be described presently. In the Stationary Simulation Study, there were 12 events in the data transformation (Table 5.1).

Table 5.1 List of interface events used for classification in the Stationary Simulation Study

Character	Interface Event
A	Pan up
V	Pan down
<	Pan left
>	Pan right
-	Zoom out
+	Zoom in
x	Center zoom
*	Impossible pan or zoom command
b	Zoom in one level through zoom slider
B	Zoom out one level though zoom slider
C	Zoom out two levels through zoom slider
R	Reset map

All events were map operations related to zooming, panning, centering and resetting the view. Detailed discussions of the operations are presented in section 4.2. Not all possible map operations were represented, because the participants did not utilize all affordances in the interface. In particular, the zoom slider allowed five levels of zoom in and zoom out, for a total of ten zoom slider operations, but only three of the ten were encountered in the course of the experiment.

In the Field and Virtual Reality Study (Table 5.2), additional event symbols denoted a switch to each of six target addresses. This difference in event sets between the experiments was due to the freedom to select targets freely in one experiment but not the other. Participants in the Field and Virtual Reality chose to utilize fifteen of the available map operations, as opposed to choosing twelve of the available operations for the Stationary Simulation Study. Of the ten possible zoom-slider operations mentioned in the previous paragraph, participants utilized six distinct operations, while in the Stationary Simulation Study they had utilized three. Rounding up the set of features was a categorical variable denoting whether the participant had worked in the field or virtual environment (Table 5.2 lists all interface classification features).

Table 5.2 List of interface events used for classification in the Field and Virtual Reality Study

Character	Interface Event
A	Pan up
V	Pan down
<	Pan left
>	Pan right
-	Zoom out
+	Zoom in
x	Center zoom
*	Impossible pan or zoom command
b	Zoom in one level through zoom slider
c	Zoom in two levels through zoom slider
e	Zoom in four levels through zoom slider
B	Zoom out one level though zoom slider
C	Zoom out two levels through zoom slider
D	Zoom out three levels through zoom slider
J	Change target to address 1
K	Change target to address 2
L	Change target to address 3
M	Change target to address 4
N	Change target to address 5
P	Change target to address 6
Z	Change target
R	Reset map

Even though the second set of operations appears considerably larger than the first, conceptually the two are near-identical. The extra zoom actions in the Field-and-Virtual-Reality group pertain to the same widget, the zoom slider, which was unchanged from the first experiment, but was used more by participants in the second experiment. The core difference between the action sets was the addition of a change-address functionality, which added a degree of freedom to how users could approach the task.

The divergence between the two operation sets was evidence for the broader relevance of our detection approach, as it was proved successful in both cases.

5.1.2 Geo-tagging and location attributes for interface actions

In the Virtual Reality Study, participants were not stationary – they moved through a "physical" (real or virtual) environment and continually changed their geographical location. The motion of participants in virtual or real space enabled the geo-tagging of interface commands. All interface events had GPS coordinates, which reflected the "ground" location at which the interface event occurred, as opposed to the software map location of the event. Consequently, the availability of geotagged information allowed reasoning about the location of user behaviors. Actions were viewed within concentric circles centered on target addresses as the focal points (Figure 5.1).

Geo-tagging allowed us to view interface actions that were initiated within an arbitrary radius (e.g. 30 m, 40 m, 50 m, 60 m) of each of the six addresses. When statistical analysis and machine-learning schemes were applied to geo-tagged interface events, a Radius parameter described the physical (or virtual for the Field and Virtual Reality Study) area of the map where the behaviors occurred. For example, a radius of 60 m in Table 5.6 means that the results of automatic detection were based on the set of interface events that occurred within 50 meters of each of the six target addresses. In Figure 5.1, a radius of 60 m corresponds to the third ring around an address. Chapter 6 includes a survey of statistically significant software differentials in the Field and Virtual Reality Study. In contrast, the stationary simulation study did not allow user movement and could not be viewed from a geo-tagged perspective.

Chapter 6 is going to show in more detail that behavioral differentials happened outside of the immediate vicinity of the target, which corresponds to the "Approach Phase" of the decision models in Chapter 4. The decision models from the Paper Map Study were characterized by multiple differentials in the approach phase and few differentials in the verification phase. The geo-located user actions in the Field and Virtual Reality study mirrored this pattern. This correspondence between the Paper Map and Field and Virtual Reality studies furnishes extra evidence that statistical outcomes from the two software experiments are not flukes, but rather indicators of a systemic link between user ability and behavior.

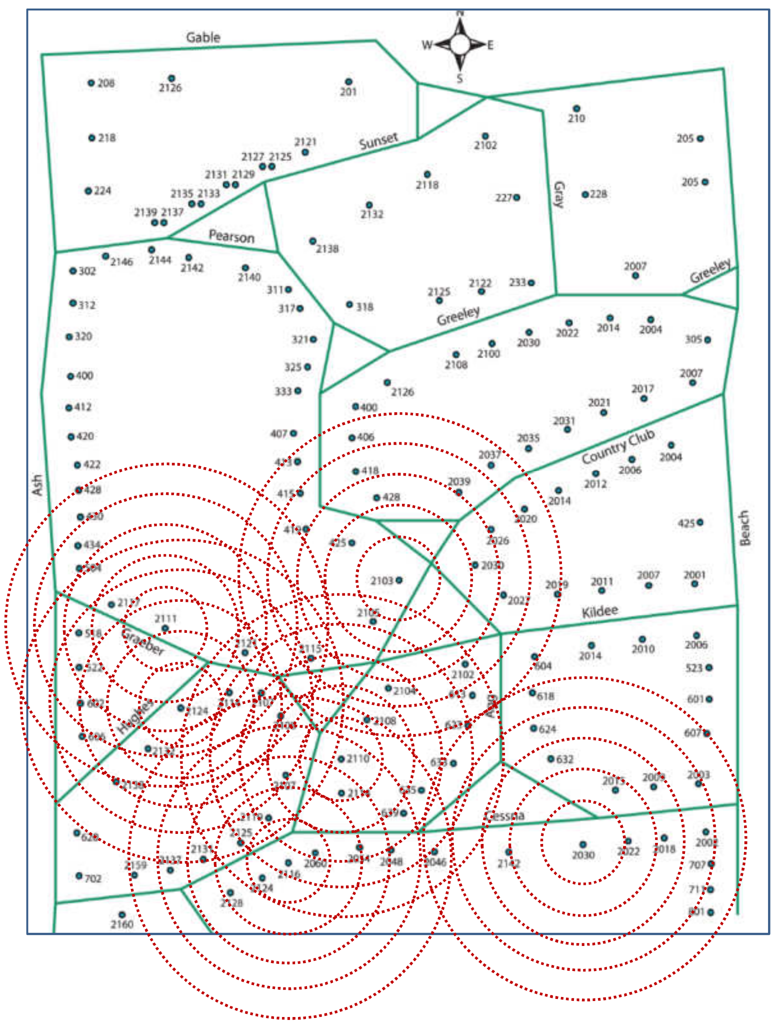


Figure 5.1 A visualization of concentric circles of differing radii centered on the six target addresses of the Field and Virtual Reality study. Geo-tagged interface events inside the areas of the circles were used for spatial visualization detection. The outermost circles have radii of approximately 100 m.

5.1.3 Data representation

The detection-ready data representation was in ARFF format. This format consists of a collection of observations together with a collection of attributes. Each feature was an interface action that had been encountered in at least one user. The occurrences of each action in each user's history were counted. To illustrate the process, Figure 5.2 presents an ARFF file for the Stationary Simulation Study. Each user is assigned one of two possible labels, "high-spatial" or "low-spatial".

```
% 1. Title: ARFF file for Stationary Simulation Study with
%    twelve attributes and four participants.

@relation '1-grams-weka.filters.unsupervised.attribute.Remove-R2-3,7-
8,11-weka.filters.unsupervised.attribute.Remove-R1,14-15,17'

@attribute + numeric
@attribute x numeric
@attribute > numeric
@attribute A numeric
@attribute < numeric
@attribute V numeric
@attribute - numeric
@attribute C numeric
@attribute b numeric
@attribute R numeric
@attribute B numeric
@attribute * numeric
@attribute vz_cat {Low,High}

@data
26,9,10,4,6,4,3,0,0,0,0,0,High
25,14,10,30,13,29,5,0,0,0,0,0,High
22,16,5,10,6,16,3,1,1,0,0,0,High
... (21 participants omitted)
20,28,15,14,11,36,3,0,0,4,11,0,Low
```

Figure 5.2 An ARFF file with twelve features, four participants and a classification attribute (vz_cat = Spatial Visualization Category). Each *@attribute* line represents a feature that participants are categorized on. Each line past the *@data* tag represents a single participant, and is a collection of numeric values. The WEKA framework parses this type of file and enables the execution of machine learning algorithms.

As seen in Figure 5.2, the final form of the data is a set of numeric values associated with each participant. Given these collections of numbers, machine learning schemes attempted to detect visualization ability through a variety of techniques, from hyperplanes to Bayesian inference to decision trees. The results sections will show that meta-classifiers were the most effective at separating high-spatial-visualization from low-spatial-visualization participants.

5.1.4 Algorithms

The data in ARFF format was processed by a battery of 52 stock machine learning algorithms from the Weka 3.7.3 framework. The algorithms represented as many methodological groups of machine learning approaches as could be obtained through the Weka framework without installing additional software. The groups in the battery included Naïve Bayes, nearest-neighbor, meta, rules-based, tree-based, neural-network-based, regression-based, support-vector-machine, and miscellaneous algorithm types. The complete list of algorithms appears in Figure 5.3. The parameters for the algorithms used in our work were not changed from their default settings in the Weka framework. Although there were many tweakable algorithm settings, and consequently an inexhaustible variety of learning schemes, if user behavior were systematically linked to spatial visualization ability, several algorithms ought to have succeeded. However, the literature search in Chapter 2 had revealed that detection of cognition-related variables is at an early stage of exploration, so spatial visualization detection was by no means guaranteed.

```
bayes.BayesianLogisticRegression -D -Tl 5.0E-4 -S 0.5 -H 1 -V 0.27 -R R:0.01-316
bayes.BayesNet -D -Q bayes.net.search.local.K2 -- -P 1 -S BAYES -E
        bayes.net.estimate.SimpleEstimator -- -A 0.5
bayes.NaiveBayes
functions.Logistic -R 1.0E-8 -M -1
functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
functions.RBFNetwork -B 2 -S 1 -R 1.0E-8 -M -1 -W 0.1
functions.SimpleLogistic -I 0 -M 500 -H 50 -W 0.0
functions.SMO -C 1.0 -L 0.0010 -P 1.0E-12 -N 0 -V -1 -W 1 -K
        functions.supportVector.PolyKernel -C 250007 -E 1.0
functions.SPegasos -F 0 -L 1.0E-4 -E 500
functions.VotedPerceptron -I 1 -E 1.0 -S 1 -M 10000
lazy.IB1
lazy.IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A
        weka.core.EuclideanDistance -R first-last"
lazy.KStar -B 20 -M a
lazy.LWL -U 0 -K -1 -A "weka.core.neighboursearch.LinearNNSearch -A”
        weka.core.EuclideanDistance -R first-last" -W trees.DecisionStump
meta.AdaBoostM1 -P 100 -S 1 -I 10 -W trees.DecisionStump
meta.Bagging -P 100 -S 1 -num-slots 1 -I 10 -W trees.REPTree -- -M 2 -V 0.0010 -N 3 -S
        1 -L -1
meta.Dagging -F 10 -S 1 -W functions.SMO -- -C 1.0 -L 0.0010 -P 1.0E-12 -N 0 -V -1 -W 1
        -K "functions.supportVector.PolyKernel -C 250007 -E 1.0"
meta.Decorate -E 10 -R 1.0 -S 1 -I 10 -W weka.classifiers.trees.J48 -- -C 0.25 -M 2
meta.ClassificationViaRegression -W weka.classifiers.trees.M5P -- -M 4.0
meta.END -S 1 -I 10 -W meta.nestedDichotomies.ND -- -S 1 -W trees.J48 -- -C 0.25 -M 2
meta.FilteredClassifier -F "supervised.attribute.Discretize -R first-last" -W trees.J48
        -- -C 0.25 -M 2
meta.LogitBoost -P 100 -F 0 -R 1 -L -1.7976931348623157E308 -H 1.0 -S 1 -I 10 -W
        trees.DecisionStump
meta.MultiBoostAB -C 3 -P 100 -S 1 -I 10 -W weka.classifiers.trees.DecisionStump
meta.RealAdaBoost -P 100 -H 1.0 -S 1 -I 10 -W weka.classifiers.trees.DecisionStump
meta.RandomCommittee -S 1 -num-slots 1 -I 10 -W weka.classifiers.trees.RandomTree -- -K
        0 -M 1.0 -S 1
meta.RandomSubSpace -P 0.5 -S 1 -num-slots 1 -I 10 -W weka.classifiers.trees.REPTree --
        -M 2 -V 0.0010 -N 3 -S 1 -L -1
meta.Stacking -X 10 -M "rules.ZeroR " -S 1 -num-slots 1 -B "rules.ZeroR"
misc.HyperPipes
misc.VFI -B 0.6
rules.ConjunctiveRule -N 3 -M 2.0 -P -1 -S 1
rules.DecisionTable -X 1 -S "BestFirst -D 1 -N 5"
rules.DTNB -X 1
rules.FURIA -F 3 -N 2.0 -O 2 -S 1 -p 0 -s 0
rules.JRip -F 3 -N 2.0 -O 2 -S 1
rules.NNge -G 5 -I 5
rules.OLM -R 0 -C 1 -U 0
rules.OneR -B 6
rules.PART -M 2 -C 0.25 -Q 1
rules.Ridor -F 3 -S 1 -N 2.0
trees.ADTree -B 10 -E -3
trees.BFTree -S 1 -M 2 -N 5 -C 1.0 -P POSTPRUNED
trees.DecisionStump
trees.FT -I 15 -F 0 -M 15 -W 0.0
trees.J48 -C 0.25 -M 2
trees.J48graft -C 0.25 -M 2
trees.LADTree -B 10
trees.LMT -I -1 -M 15 -W 0.0
trees.NBTree
trees.RandomForest -I 10 -K 0 -S 1
trees.RandomTree -K 0 -M 1.0 -S 1
trees.REPTree -M 2 -V 0.0010 -N 3 -S 1 -L -1
trees.SimpleCart -S 1 -M 2.0 -N 5 -C 1.0
```

Figure 5.3 The Weka 3.7.3 framework enabled the application of a battery of 52 machine learning algorithms to individual-user interface event sequences. The algorithms were run with their default settings in the framework, included here for verification purposes. The conceptual group of each algorithm is also displayed, and includes Naïve Bayes, nearest-neighbor, meta, rules-based, tree-based, neural-network-based, regression-based, support-vector-machine, and miscellaneous algorithms.

5.2 Stationary Simulation Results

5.2.1 Participant demographics

Twenty-five individual-user software logs were obtained from the Stationary Simulation Study. Demographically, the group was comprised of 14 females and 11 males. Twelve participants (the low-spatial-visualization subgroup) had scored 10 or fewer points on the VZ-2 measure (Ekstrom et al, 1976), while 13 participants (the high-spatial-visualization subgroup) had scored 14.75 points or more (Table 5.3).

Table 5.3 Basic characteristics of users in the Stationary Simulation Study. Low-spatial-visualization participants scored 10 points or less on the VZ-2 test (Ekstrom 1976). Conversely, high-spatial-visualization participants scored 14.75 points or more.

	Low-spatial-visualization	High-spatial-visualization	Total
Women	7	7	**14**
Men	5	6	**11**
Total	**12**	**13**	

5.2.2 Results

Classification accuracy (percentage of correct guesses) was measured for 25-fold leave-one-out classification. The baseline of 52% correct classification was obtained by the computer predicting that every participant belonged the high spatial visualization group, which was larger, with 13 out of 25 participants. Four algorithms performed with accuracy of 80% or greater, and twelve more algorithms performed with accuracy between 70% and 80%. The two best-performing algorithms predicted spatial visualization ability correctly in 84% of the cases. Table 5.4 contains the detection results.

Table 5.4 Classification success of spatial visualization ability in the Stationary Simulation Study. The baseline of 52% correct classification was achieved by the computer predicting that every participant had high spatial visualization ability. The two best-performing algorithms detected spatial visualization ability correctly in 84% of the cases.

Algorithm	Number Correct	Number Incorrect	Number Lows Incorrect	Number Highs Incorrect	Accuracy (percent correct)
Support Vector Machine	21	4	4	0	84%
Bagging with REPTree	21	4	3	1	84%
Decorate with J48	20	5	3	2	80%
Naïve Bayes	20	5	3	2	80%
ClassificationViaRegression with M5`	19	6	4	2	76%
IB1	19	6	3	3	76%
IBk	19	6	3	3	76%
MultiBoostAB with Decision Stump	19	6	5	1	76%
RealADABoost	19	6	4	2	76%
LAD Tree	18	7	4	3	72%
Logistic Model Tree	18	7	5	2	72%
RandomSubSpace with RepTree	18	7	4	3	72%
Random Tree	18	7	4	3	72%
RBFNetwork	18	7	4	3	72%
Simple Logistic	18	7	5	2	72%
Voting Feature Intervals	18	7	3	4	72%
Baseline (predict high)	13	12	12	0	52%

From the above outcomes, we concluded that automatic detection in the Stationary Simulation Study surpassed the 80% accuracy threshold that we set out as the goal of the dissertation. A range of algorithms (16 out of 52) performed considerably better than the baseline, supporting the notion that individual differences manifest systematically in interface behaviors.

The results of the Stationary Simulation Study suggested that a constrained address verification protocol lends itself to automatic detection of spatial visualization ability. Constraints in the Stationary Simulation Study were as follows: (a) no movement was afforded to participants; (b) a fixed set of ground cues appearing on two computer screens were pre-selected by experimenters; and (c) the software workflow enforced that each address be engaged in a particular order and by following a particular set of verification steps. To establish the validity of detection results beyond the borders of the restricted study protocol, we needed a separate experiment without the constraints of the Stationary Simulation study's design. That role was filled by the Field and Virtual Reality experiment, where participants were allowed freedom of movement, freedom of cue acquisition from reality, and freedom to verify addresses in arbitrary order and as many times as they desired.

5.3 Field and Virtual Reality Results

As mentioned in the previous section, successful detection in the Stationary Simulation Study fulfilled the expectation that spatial visualization ability is automatically detectable in constrained-protocol and constrained-environment scenarios. To strengthen the position of this dissertation, we now proceed with results from the comparatively unconstrained Field and Virtual Reality Study.

5.3.1 Participant demographics

Thirty-one individual-user software logs were obtained from the Field and Virtual Reality Study. Demographically, the group was comprised of eighteen females and thirteen males. Seventeen participants (the low-spatial-visualization subgroup) had scored less than 9 points on the VZ-2 measure (Ekstrom et al, 1976), while fourteen participants (the high-spatial-visualization subgroup) had scored 15 points or more (Table 5.5).

Table 5.5 Basic characteristics of users in the Field and Virtual Reality Study. Low-spatial-visualization participants scored less than 9 points on the VZ-2 test (Ekstrom 1976). Conversely, high-spatial-visualization participants scored 15 points or more.

	Low-spatial-visualization	High-spatial-visualization	Total
Women	11	7	**18**
Men	6	7	**13**
Total	**17**	**14**	

5.3.2 Results: Field and Virtual Reality Study

Detection accuracy in percent was measured for 31-fold leave-one-out classification. The accuracy baseline (classifying every participant as low-spatial-visualization, which would be correct in 17 out of 31 cases) was 54.84%. Four detection algorithms performed with greater than 80% accuracy, and seventeen more algorithms performed with accuracy between 70% and 80%. The two best-performing algorithms detected spatial visualization ability correctly in 87.10% of the cases. Table 5.6 shows detection outcomes.

The high accuracy of automatic detection in the Field and Virtual Reality Study validated the story that had been anticipated by the Paper Map Study and corroborated by the Stationary Simulation Study: systematic differences in behavior at the interface reveal the spatial visualization ability of the user. Algorithm accuracy in both the constrained-protocol experiment and the relaxed-protocol experiment has established an initial level of ecological validity of automatic detection.

5.4 Interpretation of Algorithmic Outcomes

This section investigates the type and mechanics of algorithms that correctly determined participants' ability. A closer look at the prediction models will verify that detection was meaningful and will connect the results to statistics from previous chapters.

5.4.1 Plurality of algorithms

Before examining the top performers, the issue of detection quality versus quantity needs to be addressed. A plurality of algorithms with relatively high accuracy is an important adjunct to the quality of detection. If the results indicated a few successful algorithms while the rest were insignificant, the validity of the detection claim would be diminished. Conversely, a large quantity of relatively good performers supplementing the top performers implies that behavioral differences are systematic and discoverable by diverse approaches.

Table 5.6 Automated detection success on the Stationary Simulation Study. The baseline of 54.84% correct classification was achieved by the computer predicting that every participant had low spatial visualization ability. The two best-performing algorithms detected spatial visualization ability correctly in 87.10% of the cases. The Radius column describes the area from which geo-tagged interface events were taken. The area was comprised of six circles centered on the six target addresses. For example, a radius of 65m means that interface actions that occurred within 65 m of each of the six addresses were used for detection.

Radius from target	Algorithm	Number Correct	Number Incorrect	Number Lows Incorrect	Number Highs Incorrect	Accuracy (percent correct)
65m	ClassificationViaRegression with M5`	27	4	2	2	87.10%
70m	Naïve Bayes Tree	27	4	2	2	87.10%
55m	BFTree	25	6	2	4	80.65%
95m	FURIA	25	6	0	6	80.65%
95m	ADTree	24	7	3	4	77.42%
60m	Bagging with REPTree	24	7	5	2	77.42%
50m	DecisionStump	24	7	5	2	77.42%
95m	JRip	24	7	2	5	77.42%
50m	Locally Weighted Learning	24	7	2	5	77.42%
90m	LogitBoost	24	7	3	4	77.42%
80m	OneR	24	7	2	5	77.42%
70m	RandomForest	24	7	3	4	77.42%
65m	Conjunctive Rule	23	8	1	7	74.19%
65m	IB1	23	8	5	3	74.19%
65m	IBk	23	8	5	3	74.19%
65m	RandomCommittee with RandomTree	23	8	2	6	74.19%
60m	RealAdaBoost with DecisionStump	23	8	4	4	74.19%
65m	RIDOR	23	8	4	4	74.19%
65m	Decorate with J48	22	9	5	4	70.97%
50m	MultiBoostAB with DecisionStump	22	9	7	2	70.97%
65m	Support Vector Machine	22	9	4	5	70.97%
105m	SimpleCart	22	9	7	2	70.97%
N/A	Baseline (predict low)	17	14	0	15	54.84%

The results from both experiments show that considerable portions of the algorithm battery were moderately successful or better. Sixteen out of 52 algorithms for the Stationary Simulation Study and 21 out of 52 algorithms on the Field and Virtual Reality experiment exhibited better than 70% accuracy against respective baselines of 52% and 54.84%. Further, eight algorithms overlapped between studies: Bagging, Support Vector Machine, Decorate, RealAdaBoost, MultiBoostAB, IB1, IBk, and ClassificationViaRegression. Five of the eight belong to the meta family of algorithms, where a base algorithm is informed and improved upon by the meta-algorithm. This outcome is consistent with theoretical results in the machine learning literature that point to "meta" algorithms reducing generalization error compared to base algorithms (e.g. Krogh and Vedelsby (1995) for ensembles, and Wolpert (1992) for stacked generalizers). Meta-algorithms were used with the exact same default parameters in both experiments, and no custom tailoring of algorithms was performed.

5.4.2 Analysis of the best detection schemes

Eight algorithms overcame the 80% threshold set in this dissertation. Of those eight, two algorithms achieved 87.10% accuracy on the Field and Virtual Reality Study and two more achieved 84% accuracy on the Stationary Simulation Study, for a total of four top-performing algorithms. One algorithm from each pair performed well only on one study, and one algorithm from each pair performed well in both studies. Therefore, the two absolute best performers were (a) ClassificationViaRegression, with accuracy of 87.10% in the Field and Virtual Reality experiment and 76% in the Stationary Simulation Study and, and (b) Bagging with accuracy of 84% in the Stationary Simulation Study and 77.42% in the Field and Virtual Reality Study. Both winner algorithms were meta-algorithms that worked particularly well with small samples and unstable distributions of observations. The following paragraphs examine the prediction models constructed by the two best algorithms.

5.4.2.1 Prediction models for the Stationary Simulation Study

The BAGGing (Bootstrap AGGregating) machine learning approach was proposed by Breiman (1996). It is an ensemble classification approach, which means that it aggregates the votes of multiple classifiers to reach a decision. The central technique of the algorithm is to create multiple learning datasets from the original learning data set by drawing data points at random *with replacement.* Creating a dataset by drawing with replacement is called bootstrapping. As a result, in each of these new “bootstrap” data sets, the same data point can be present multiple times. If the size of the new data set is equal to the size of the original data set, the probability of a particular data point appearing at least once is approximately 0.632 (Breiman 1996, p. 136).

After the bootstrap samples have been created, a classifier is built from each new learning set, and then all the classifiers vote on the original data. The Weka software default instructions were to create 10 classifiers of type REPTree. REPTree is a binary-decision-tree classifier that splits the feature space consecutively at the point that minimizes misclassification error, effectively creating hyperrectangles, each labeled with a particular class, that cover the feature space. The ten decision trees for the Stationary Simulation Study are drawn in Figure 5.4 on the next page.

Figure 5.4 exhibits ten different decision trees that arose from ten bootstrap samples created by drawing from the original sample with replacement. Zoom-in, zoom-out actions, pan-up, pan-down, center-zoom, and reset-map actions are the driving features in this classification. More map operations of any kind usually let the classifier decide that the participant is of low-spatial-visualization ability. The lone exception is Tree 4, where the algorithm decided high-spatial-ability participants inhabit the region between 22 and 25.5 zoom-in operations. But even that classification tree had a higher-priority rule assigning low spatial visualization ability to participants who executed five or more pan-up operations. Panning was the most-heavily-utilized feature for detection, followed by zooming, centering the zoom, and map resets. The decisions of the Bagging algorithm are human-interpretable and congruent with the behavioral expectations built by our traditional statistical analyses in Chapters 3 and 4.

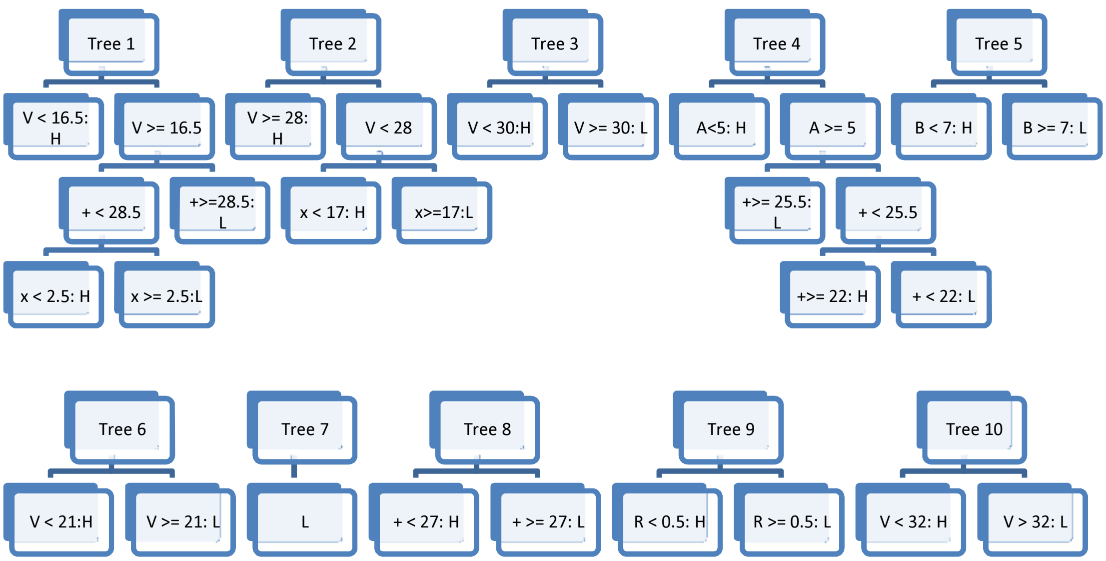


Figure 5.4 The ten REP tree classifiers created by the Bagging algorithm for the Stationary Simulation Study. H denotes High spatial visualization ability; L denotes Low spatial visualization ability; V is the pan-down action; A is the pan-up action; B is an action zooming out one level through the slider widget; + is the zoom-in action; x is the center-zoom action; R is the reset-map action. In general, more map actions are associated with lower spatial visualization ability. Trees are quite distinct from one another due to the sampling process producing variegated bootstrap samples.

The other algorithm with best overall performance on both experiments was the meta classifier ClassificationViaRegression (Frank, Wang, Inglis, Holmes, and Witten 1998) which implements a M5` tree (Quinlan 1992, Wang and Witten 1997). The M5` algorithm makes binary decisions that partition the instance space to minimize the mean squared error between the model’s predictions and the class labels of 0 and 1 (low spatial visualization and high spatial visualization ability). When the tree is pruned, the leaves become linear regression models that contain the attributes in the pruned subtrees. A classifying tree with regressions at the leaves is built for each class label (low spatial visualization ability and high spatial visualization ability).

For the Stationary Simulation study, the ClassificationViaRegression algorithm produced two trivial (single-node) M5 trees that reduce to two linear regression models:

Low spatial visualization score:

(1) $Y_L = -0.2282 + 0.0149 * center\_zoom + 0.229 * pan\_down$

High spatial visualization score:

(2) $Y_H = 1.2282 - 0.0149 * center\_zoom - 0.229 * pan\_down$

During detection, the algorithm computes both Y_L and Y_H values and chooses the class label with the higher score. In linear model (1), more pan-down and center-zoom actions contributed to a higher score for low spatial visualization ability, while in linear model (2), more pan-down and center-zoom actions decreased the score for high spatial visualization ability. This combination of linear models acted on a subset of the classification features utilized by the Bagging algorithm, and it only contained two instead of ten voting classifiers. The relative simplicity of the classification scheme may explain ClassificationViaRegression's lower detection accuracy of 76% against the 84% achieved by the Bagging classifier.

5.4.2.2 Prediction models for the Field and Virtual Reality Study

For the Field and Virtual Reality Study, the Bagging algorithm produced the following ten classification trees (shown in Figure 5.5). The salient classification features in Figure 5.5 are zoom-in actions, pan-up actions, pan-left actions, pan and zoom actions that were impossible, and target address switches. As with the Stationary Simulation experiment, more map operations were related to lower spatial visualization ability. More target address switches were a sign of less robust planning or difficulties with address completion. Impossible zoom and pan actions also indicated difficulties with the task. The extra margin of freedom in the interface of the Field and Virtual Reality experiment allowed non-map operations, in this case, intention-signaling (through target switching) to became a new source of detection.

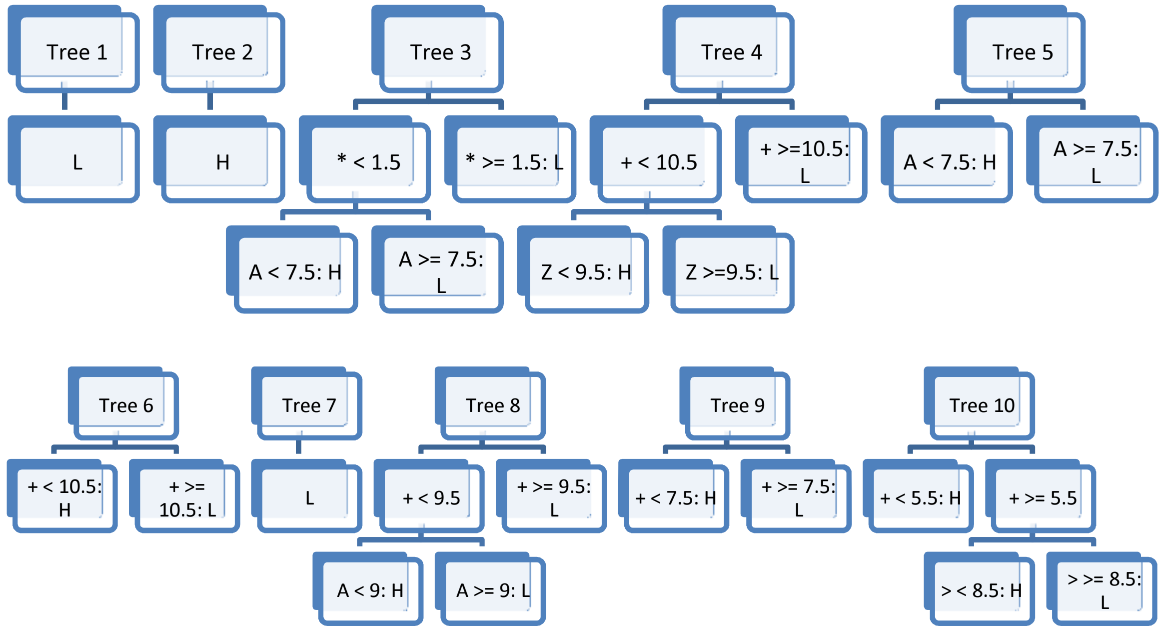


Figure 5.5. The ten REP tree classifiers created by the Bagging algorithm for the Field and Virtual Reality Study. H denotes High spatial visualization ability; L denotes Low spatial visualization ability; > is the pan-left action; A is the pan-up action; Z is the action to switch the target address; + is the zoom-in action; * is a pan or zoom action that was not possible on the map. Like in the Stationary Simulation Study, more map actions are associated with lower spatial visualization ability. Zoom actions are the most salient classification feature, followed by pans. Impossible pan or zoom commands and target switching also signal differences in spatial visualization ability.

The other top-performing detection scheme, ClassificationViaRegression, reached an accuracy of 87.10% on the Field and Virtual Reality experiment through the following set of models (Figure 5.6).

The ClassificationViaRegression detection scheme operated on a similar set of map operations as the Bagging scheme, selecting the two most salient features from Bagging. However, the M5P decision tree distinguished cases where map operations had differing relative weights in determining the predicted class, thereby highlighting sub-groups of low- and high-spatial-visualization participants. In the low-spatial-visualization model, participants with fewer pan actions were considered less likely to have low spatial visualization ability, but their zoom actions were more than twice as important as their pan actions in increasing the likelihood of low spatial visualization.

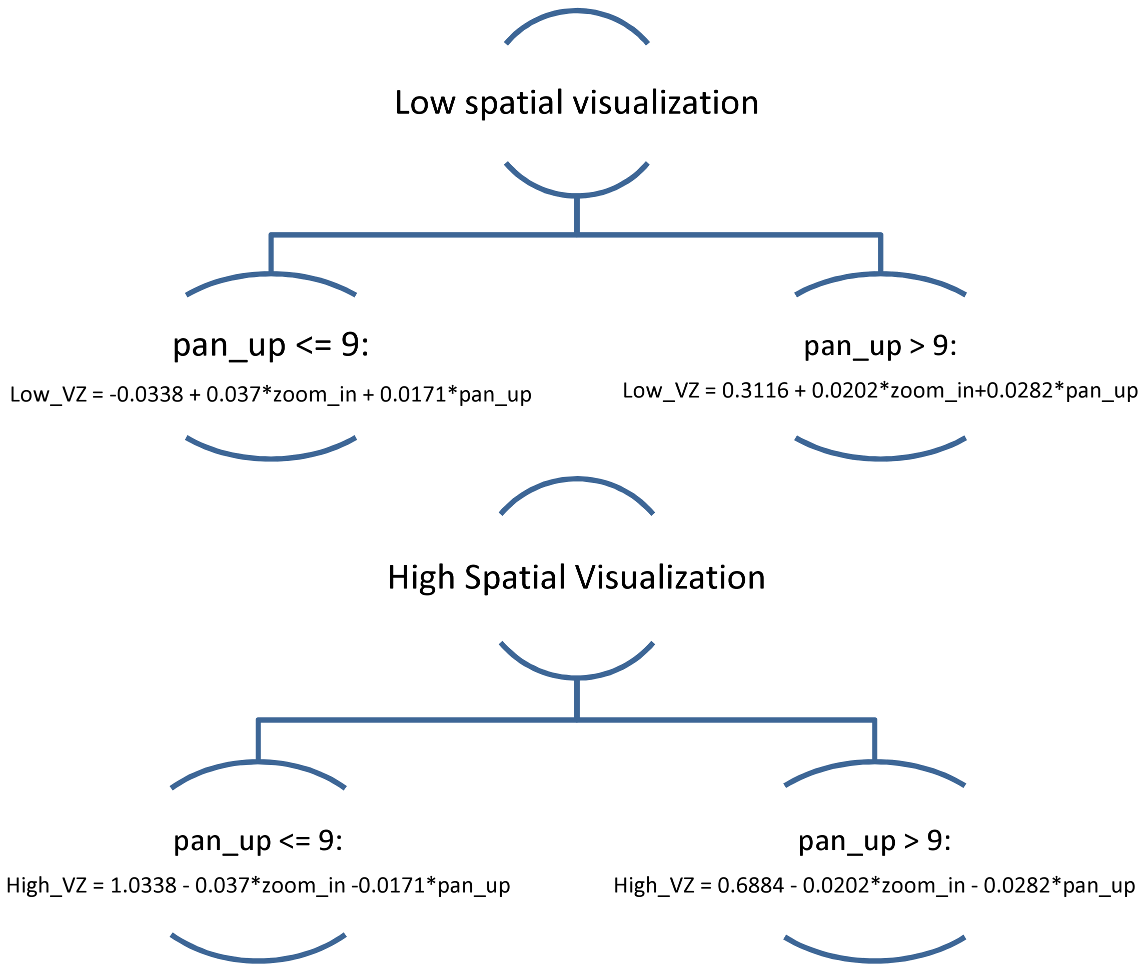


Figure 5.6. Field and Virtual Reality Study M5` classification trees with regressions at the leaves. In the low spatial visualization classifier, more map operations led to a higher score for low spatial visualization ability, while in the high spatial visualization classifier , more map operations led to a lower score for high spatial visualization ability and a lower likelihood of a high spatial visualization prediction. What is noteworthy is the classifier has isolated subgroups within each class that have differing likelihoods to be labeled low- or high-spatial-visualization, and their map operations have differing relative weights.

Conversely, participants with more pan actions were considered more likely to have low spatial visualization ability, and each extra pan action increased that likelihood even faster. In the high-spatial-visualization model, the reverse was true: extra zoom actions mattered more to participants who panned infrequently, while frequent panners were considered low-spatial-ability candidates. Overall, the complexity of each regression model allowed the

ClassificationViaRegression algorithm to separate low- and high-spatial ability cohorts into further subgroups where map operations had differing weights, and this finer level of distinction may have been the reason for the high accuracy of the scheme.

5.4.3 Conclusion

This subsection investigated the quantity, quality, and predictive models of algorithms that exhibited higher performance on the Stationary Simulation and Field and Virtual Reality studies. Four outcomes became evident. First, there was a plurality of algorithms that detected spatial visualization ability, with eight algorithms meeting the 80% detection accuracy goal (four in each study), and twenty-two more algorithms achieving rates between 70% and 80% on one or both of the studies. The long list of useful detection schemes lends strong support to the expectation that spatial visualization ability will be detectable in future applications.

Second, four algorithms achieved detection rates of 84% on the Stationary Simulation Study or 87.10% on the Field and Virtual Reality study, which means that in the pack of 30 "good" algorithms there are some that are "excellent", and therefore practicable.

Third, algorithms of the "meta" type, which is characterized by various forms of classifier aggregation and extension, were densely represented among high performers. This outcome is consistent with theoretical results in the machine learning literature that point to "meta" algorithms reducing generalization error compared to base algorithms (e.g. Krogh and Vedelsby (1995) for ensembles, and Wolpert (1992) for stacked generalizers).

Fourth, the most successful predictive models built by algorithms were intuitively meaningful. Machine learning in multidimensional spaces can fail to make sense intuitively, which leaves investigators with no assurance that the decision making was not based on meaningless patterns or flukes in the data. This was not the case in the current work. The models

(a) supplied an understandable interpretation of the decision making and (b) validated the effectiveness of data transformations and algorithms.

5.5 Results in Light of the Existing Literature

As described in more detail in Chapter 2, user-sensing results in the current literature have two or more of the following conditions: (a) investigating volatile (rather than semi-permanent) cognitive variables such as learning and frustration; (b) relying on one or more sensors such as galvanic skin response, pressure, and gaze tracking; (c) detection rates in the 70% range and no higher; (d) user variables not connected with personal computer ability; and (e), results not related to classification of future users. For example, Chang et al. (2013) achieved 98% accuracy in cognitive fingerprinting from keystroke dynamics, but did not detect user variables other than “uniqueness”. Therefore, their research was unrelated to the current dissertation.

In contrast, our results allow the detection of a semi-permanent cognitive ability that is known to relate to user’s computer performance. The detection technique does not depend on sensors and is immediately deployable on generic computers. Detection rates exceeded 80% and approached 90%. The algorithm does not need to know whether the user made mistakes, so detection can be unaware of environmental conditions, and therefore need no sensors! The final outcome is that we can detect user ability on a complex, location-based, professional task without any knowledge of the environment, considerably reducing both hardware and software costs through simplification. The dissertation result makes inroads into intelligent interfaces for professional workflows - with a shortcut completely bypassing the environment!

5.6 How to Deploy Detection

Previous sections of this chapter established quality, quantity, interpretation, and validity of detection outcomes on the Stationary Simulation and Field and Virtual Reality experiments. This section explains how the detection technique can be deployed in future projects in industry or academia.

5.6.1 Implementation methodology

In the Stationary Simulation and Field and Virtual Reality experiments, we did not observe a "one-size fits all" algorithm that achieved detection accuracy over 80% on all data sets. Instead, there were eight algorithms that had accuracy of 80% or better, but they had varying performance profiles. Four algorithms: BFTree, FURIA, Naïve Bayes, and Naïve Bayes Tree achieved high accuracy (80.00%, 80.65%, and 87.10%) on one experiment. Two algorithms, Decorate and Support Vector Machine achieved high accuracy (respectively 80% and 84%) on one experiment and accuracy of 70.97% on another. The two best performers, Bagging and ClassificationViaRegression/MP` scored 84% and 87.10% in one study and 76% and 77.42% in the other study. Table 5.7 shows comparative performance of notable algorithms on both experiments.

There are three observations that can be made about Table 5.7. First, there were no "silver-bullet" algorithms with greater than 80% performance on all data sets. Second, algorithms segregated into six performance groups with differing accuracy profiles, and five of the six groups performed considerably better in one experiment. Third, eight algorithms performed "adequately" on both data sets.

Table 5.7 Comparative performance of notable detection algorithms in the Stationary Simulation and Field and Virtual Reality experiments. The algorithms grouped along different performance profiles. Detection accuracy of 80% or higher is highlighted. Accuracy less than 70.97% is not listed. Algorithms are listed in descending order of performance.

Algorithm	Detection accuracy (% correct), Stationary Simulation Study	Detection accuracy (% correct), Field and Virtual Reality Study	Group Label
Bagging	**84.00%**	77.42%	Best Performance
ClassificationViaRegression/MP`	76.00%	**87.10%**	
Decorate	**80.00%**	70.97%	High performance and some carry-over capability
Support Vector Machine	**84.00%**	70.97%	
BFTree		**80.65%**	High performance on one experiment and no carry-over capability
Furia		**80.65%**	
Naïve Bayes	**80%**		
Naïve Bayes Tree		**87.10%**	
IB1	76.00%	74.19%	Some capability in both experiments
IBk	76.00%	74.19%	
MultiBoostAB	76.00%	70.97%	
RealAdaBoost	76.00%	74.19%	
ADTree		77.42%	Some capability in the Field and Virtual Reality experiment
Conjunctive Rule		74.19%	
Decision Stump		77.42%	
JRip		77.42%	
Locally Weighted Learning		77.42%	
Logit Boost		77.42%	
OneR		77.42%	
Random Committee		74.19%	
Random Forest		77.42%	
RIDOR		74.19%	
SimpleCart		70.97%	
LAD Tree	72.00%		Some capability in the Stationary Simulation experiment
Logistic Model Tree	72.00%		
Random Tree	72.00%		
Random Subspace	72.00%		
RBF Network	72.00%		
Simple Logistic	72.00%		
Voting Feature Intervals	72.00%		

The majority of algorithms performed better in one of the two studies. Therefore, there appears to be a connection between the details of the experimental protocol and an algorithms' suitability. We propose the following approach to a deployment. As a first step, the software designers ought to roll out a pilot study to determine what algorithms perform well. After the best schemes have been found, they ought to be integrated into the second, and full, deployment. We will next discuss the expected accuracy from such an approach.

5.6.2 Accuracy expectations for new deployments

The outcomes in Table 5.6. should alleviate fears of low accuracy in the future. In particular, eight algorithms met the 80% accuracy threshold set forth in the goal of the dissertation. Twenty-two other algorithms achieved detection rates between 70.97% and 80%, for a total of 30 algorithms with notable accuracy (against baselines of 50% and 54%). Further, eight algorithms achieved notable accuracy on both data sets in parallel (between 70.97% and 87.10%). Since the protocol differences between experiments were considerable, these accuracy results are evidence that algorithms can be ported over between address verification protocols. In the absence of a pilot rollout, the expectation for detection accuracy would be at least in the 70% range. Conversely, with a pilot rollout, the top performers are expected to exceed 80%. Overfitting is not a concern as long as a representative sample is obtained for the pilot, because our accuracy estimates are based on cross-validation techniques.

5.7 Conclusion

We obtained evidence that spatial visualization ability is systematically linked to user behaviors at the interface, and that a generic computer device can be trained to recognize user spatial visualization from map-oriented, location-based interfaces. A multitude of reports in the literature relate spatial visualization and computer usage. But how can we engineer interfaces to take advantage of visualization detection? Chapter 6 investigates data from our experiments and published work to provide answers.

CHAPTER 6. POTENTIAL SPATIO-VISUAL ADAPTATIONS

Previously, we found spatial visualization impacts performance on a location-based address verification task, and also uncovered statistically significant divergent behaviors. But how can we engineer interfaces to take advantage of visualization detection? From a system design perspective, divergent behaviors are expected to require divergent software workflows, because, in general, software workflow is tailored to user behavior. Interface adaptations are expected to maximize performance and satisfaction for individuals.

However, spatio-visual adaptation guidelines were not readily available due to the novelty of the research. The relationship between behaviors and software adaptations has not been investigated in detail. This chapter provides recommendations based on existing publications and our experimental results. The chapter first highlights leads in the literature to argue the value of spatial-visualization-based adaptation specifically. Section 6.2 draws together several reports to elicit an adaptation directive. Section 6.3 affirms the directive through additional behavioral analysis of the Field and Virtual Reality experiment. Section 6.4 presents specific adaptation recommendations. Section 6.5 concludes the discussion.

6.1 Adaptation Leads in the Literature

Literature on spatial-visualization-related software adaptations is limited due to the novelty of the research. Existing reports fall into two heavily populated and mutually exclusive categories: spatial visualization research that does not concern itself with software adaptations, e.g. Campbell (2011) compared user performance on a small-screen device and a large-screen device; or adaptation research that does not target spatial visualization, e.g. Ohm, Bienk, Kattenbeck, Ludwig, & Müller (2016) compared navigation aids for users with varying sense of direction (Figure 6.1). In contrast to the available literature, successful adaptation recommendations would require both a spatial-visualization orientation and an alternative-interface human-subject experiment.

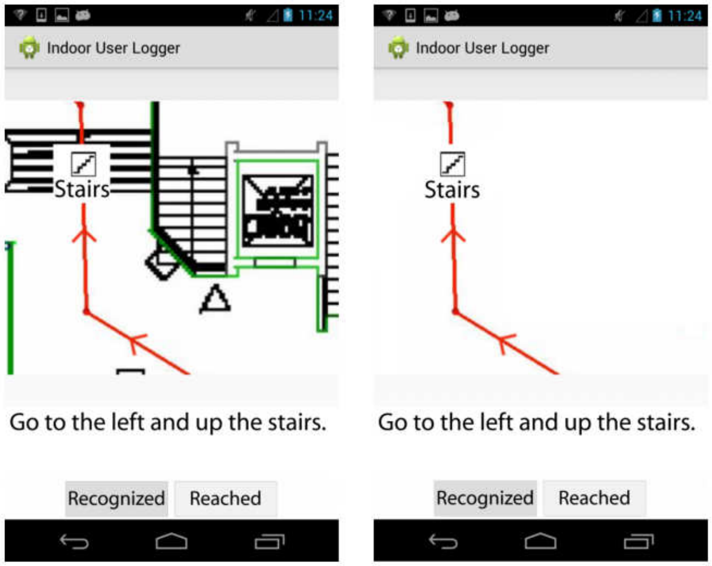


Figure 6.1 Detailed (left) and abstract indoor navigation aid in Ohm, Bienk, Kattenbeck, Ludwig, & Müller (2016). Participants with strong self-reported sense of direction performed considerably better with the abstract interface.

The lack of relevant literature can be explained by the behavioral focus of the present dissertation: existing reports are heavily weighted towards the performance characteristics of spatial visualization, and not towards how participants actually *used* the interface. In contrast, our work expected performance differentials to be a given and emphasized behavioral observation from the outset.

While glancing discussion of behavior was present in many publications, almost none can be used for adaptation discussions due to the lack of alternative-interface experiments. A notable exception is Brennan, Kelly, & Arguello (2014), who tested 21 participants of low and high spatial visualization ability (measured via Ekstrom's (1976) VZ-2 Paper Folding Test) on web search tasks in the "entertainment" and "science and technology" domains. Three types of tasks were investigated: (i) obtaining a definite answer to a question, (ii) assembling a roster of items and explanations, and (iii) generating a solution to an open-ended question. High-spatial-visualization participants performed more searches, used longer search word combinations, viewed more pages, abandoned more pages, and used the search engine result pages more, but spent less time per page. A critical finding was that user-reported workload was unaffected by spatial visualization ability (p.173), signaling that the individual difference influences user workflow subconsciously. An immediate corollary is that participants would be unable to self-select adaptations in an interface.

The behavioral results in Brennan, Kelly, & Arguello (2014) are complemented by performance results in Downing, Moore, & Brown (2005). Thirty-seven participants engaged in information retrieval via a library interface for advanced search of articles. Participants searched for articles related to two business-related questions, two biology-related questions, and a domain-neutral question used as a baseline. Spatial visualization was determined by a combined score on Ekstrom et al.'s (1976) VZ-1 Form Board Test and VZ-2 Paper Folding Test. High-spatial-visualization users found the first relevant article faster, and found more relevant articles than their low-spatial-visualization counterparts. A conclusion from this article and the spatial-visualization-related information-retrieval literature is that high-spatial-visualization users are expected to outdo other user groups during search tasks, and the performance differentials would increase with the complexity of the interface. Going a step further in explaining the phenomenon, Zhang and Salvendy (2001) posited, "*Individuals with high spatial ability, however, tend to outperform individuals with low spatial ability only when information search tasks require the use of spatial ability in mentally constructing a model of the organization and structure of embedded task information*". Since information organization is relevant to both interface design and application content, the immediate conclusion, as with Brennan, Kelly, and Arguello (2014), is that not all users could be expected to choose the best adaptations for themselves, due to a spatial visualization disadvantage.

Literature accounts of training different visualization groups describe high-spatial-visualization users as benefitting from different approaches than low-spatial-visualization users. Froese, Tory, Evans & Shrikhande (2013) showed that high- and low-spatial-visualization users benefitted differently from three types of computerized training for an orthogonal projection task. Spatial visualization ability was measured with the Vandenberg & Kuse (1978) Mental Rotations Test[3]. One-hundred-and-seventeen users were first rated on the task of choosing correct orthogonal views of three-dimensional objects. Afterwards, they received one of three types of training, which showed either static intermediate steps, or animated rotations, or no intermediate steps at all. The users then performed another set of orthogonal projection tasks, and their results were compared to their initial performance. Low-spatial-visualization participants benefitted the most from static-image training (18% improvement), less from animation-based training (17% improvement), and the least from the no-intermediate-results training (10% improvement) (p.2814). High-spatial-visualization participants had a different profile of training results. They benefitted most from no-intermediate-results training (6% improvement), then from static-intermediate-results training (5% improvement), and the least from animation-based training (1% improvement) (p. 2814). The study showed that the most effective methods of training for each group were the least effective methods for the other. The training differences reinforce an expectation that system engineers can use spatial-visualization-related adaptations to modify user performance.

[3] The MRT test is a widely-used alternative to the Ekstrom et al. (1976) Paper Folding Test.

More evidence of spatial-visualization groups being affected differentially is furnished by Nguyen (2012). Sixty low- and high-spatial-visualization participants were tested on their understanding of simple and complex anatomical objects, after being trained on (a) six canonical views of a cube; (b) six canonical views of the object; and (c) animated views of the object. Spatial visualization ability was measured with a computer version of the Vandenberg & Kuse (1978) Mental Rotations Test. Figure 6.2 shows a noteworthy outcome: while in two treatments, high-spatial-visualization participants scored higher, in the third, low-spatial-visualization participants outperformed them. After hammering on the superior computer ability of high-spatial-visualization users for most of this dissertation work, it is refreshing to see conditions in which they switch roles with their counterparts and become the underdogs. This remarkable result suggests that, under certain conditions, interface design is salient enough to dictate users' performance outcomes.

Another report of interface-dependent performance differentials will be investigated in the next section, after the adaptation guidance is stated explicitly.

6.2 A Directive for Spatio-Visual Adaptation

Despite sparse publications on spatial-visualization-based adaptation, our experimental data and several reports in the literature point to a definite distinction: high-spatial-visualization users prefer and benefit from survey knowledge, while low-spatial-visualization users prefer and benefit from landmark knowledge. This is consistent with the landmark-route-survey acquisition process described in Siegel and White (1975). In this process, a user first recognizes landmarks, then links them together to form routes, and routes can give rise to a holistic, or survey, understanding of an area. The connection between the LRS model and spatial visualization will be explained after presenting the next report.

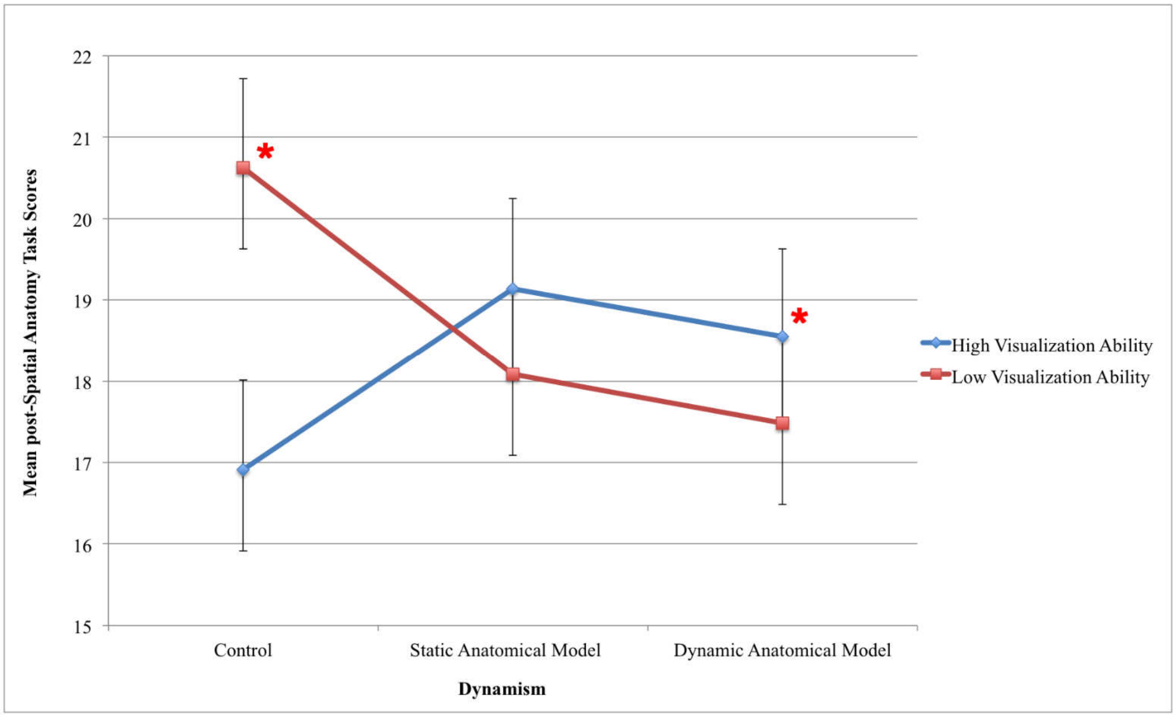


Figure 6.2. Task scores from Nguyen (2012) after three differing training modes. Low-spatial-visualization and high-spatial-visualization participants exhibit differing performance profiles, flip-flopping in performance depending on treatment.

Bay & Ziefle (2008) trained 30 participants aged 9-14 on a menu navigation task for a smart phone, with the explicit goal of observing the interaction of landmark, route, and survey knowledge and spatial visualization ability. Spatial visualization was measured through the Tewes (1983) Mosaic Test. Training was conducted in three treatments: (a) a landmark mode, where participants were given the exact menu choices to complete the task; (b) survey mode, where participants were given the entire hierarchy of all possible selections; and (c) a combined mode of landmark, survey, and route knowledge, where participants were allowed to interact with the device for five minutes. High-spatial-visualization participants performed best after pure-survey training, while low-spatial-visualization participants performed best after pure-landmark training, outperforming high-spatial-visualization participants in time, number of steps taken, and number of undo actions.

Both Bay & Ziefle (2008) and Nguyen (2012) reported training modes that set low-spatial-visualization participants ahead of their high-ability counterparts. These findings were counterintuitive and striking in light of the existing literature: despite high-spatial-visualization users being the "favorites" on computer tasks, presentation modes exist that can overturn expectations of performance. Of course, based on Williges, Elkerton, Vicente, & Hayes (1990), the impact of such designs depends heavily on task type:

> *However, just because the individual differences have been assayed and isolated does not guarantee that the accommodation will be successful. This difficulty is acknowledged by Egan and Gomez (1985, p. 215): "The step of accommodating individual differences not only tests the analyses that precede it, but it also tests the theory of how an experimental manipulation ... will change the original task." (p. c-23)*

Bay & Ziefle's (2008) dichotomy of survey and landmark knowledge preferences mapping to high- and low-spatial-visualization participants were corroborated by our own data, which will be examined in the next section. But what is the nature of the relation between landmark preference, survey preference, and spatial visualization ability? The answers can be pieced together from the accounts in Rodes & Gugerty (2012) and Meneghetti, Gyselinck, Pazzaglia, & De Beni (2009).

Rodes & Gugerty (2012) investigated sixteen participants drawing a map from memory, after having used simulated aerial navigation software for an unmanned aerial vehicle. Spatial visualization ability was tested through Ekstrom et al.'s (1976) VZ-2 Paper Folding Test. After controlling for visual memory, spatial visualization ability was significantly associated with map draw error and therefore quality of recall! This outcome was surprising and counter-intuitive, as it suggested spatial visualization had a separate effect from visual memory on the construction and retention of survey knowledge. The marginal effect of spatial visualization ability could explain why low-spatial-ability participants are less comfortable with survey knowledge and prefer landmark knowledge.

Meneghetti, Gyselinck, Pazzaglia, & De Beni (2009) ran a psychological study of 76 participants where recall of spatial and non-spatial text descriptions was measured while being interfered with through secondary tasks of spatial tapping and articulatory suppression. Spatial tapping consisted of tapping the four corners of a 30 x 24 cm rectangular board and interfered with visuospatial working memory, while the articulatory suppression task (repeating the syllables “ba-be-bi-bo-bu”) interfered with verbal working memory. Spatial visualization ability was measured by the Vandenberg & Kuse (1978) Mental Rotations Test. High-spatial-visualization participants were able to overcome the interference for the spatial text description (but not for the non-spatial description), while low-spatial-visualization participants suffered recall degradation for all treatments. These outcomes showed that spatial ability is used as an additional resource when processing spatial descriptions, and allowed high-spatial-visualization users to not require additional “executive resources”. As a result, high-VZ users appear to have extra capacity to manipulate and exploit survey knowledge that is subconscious, per the outcomes from Brennan, Kelly, and Arguello (2014). In effect, high-VZ users appear to be have a preference for survey knowledge due to a modest comic-book-hero “superpower”: they can exploit survey knowledge as it arrives without engaging additional “executive resources”. On the other hand, low-spatial-visualization participants lack the spatial-visualization “superpower” and cannot exploit survey knowledge, instead preferring landmark knowledge.

Rhodes & Gugerty (2012) and Meneghetti, Gyselinck, Pazzaglia, & De Beni (2009) appear to have clarified how the landmark-route-survey process in Siegel and White (1975) manifests itself in low- and high-spatial-visualization interface preferences. This view was corroborated by our data, as described in the next section.

6.3 Adaptation Indicators in the Field and Virtual Reality Study

To gain insight into participants' interface preferences, we further investigated their behaviors in the Field and Virtual Reality Study. For this purpose, behaviors were defined as sequences of user actions of length up to 9. Longer sequences were not considered due to computational cost. Additionally, the longer the set of actions, the less probable its replication within the set of 31 users. A total of more than 16,000 unique sequences were present in the raw data of the Field and Virtual Reality Study alone, representing a gamut of behaviors of different length. Wilcoxon-Mann-Whitney tests were performed to find statistically significant divergent behaviors consisting of up to 9 consecutive interface commands. Approximately 17,000 interface action sequences of length 1-through-9 were compared between spatial-visualization groups. Figure 6.3 presents differential user behaviors in a location-based context.

The longest behavior sequences that were statistically significant consisted of 4 actions. Only the longest sequences in a series were recorded. If subsequences were also significant at a different *p*-value, they were even more common, and were also recorded.

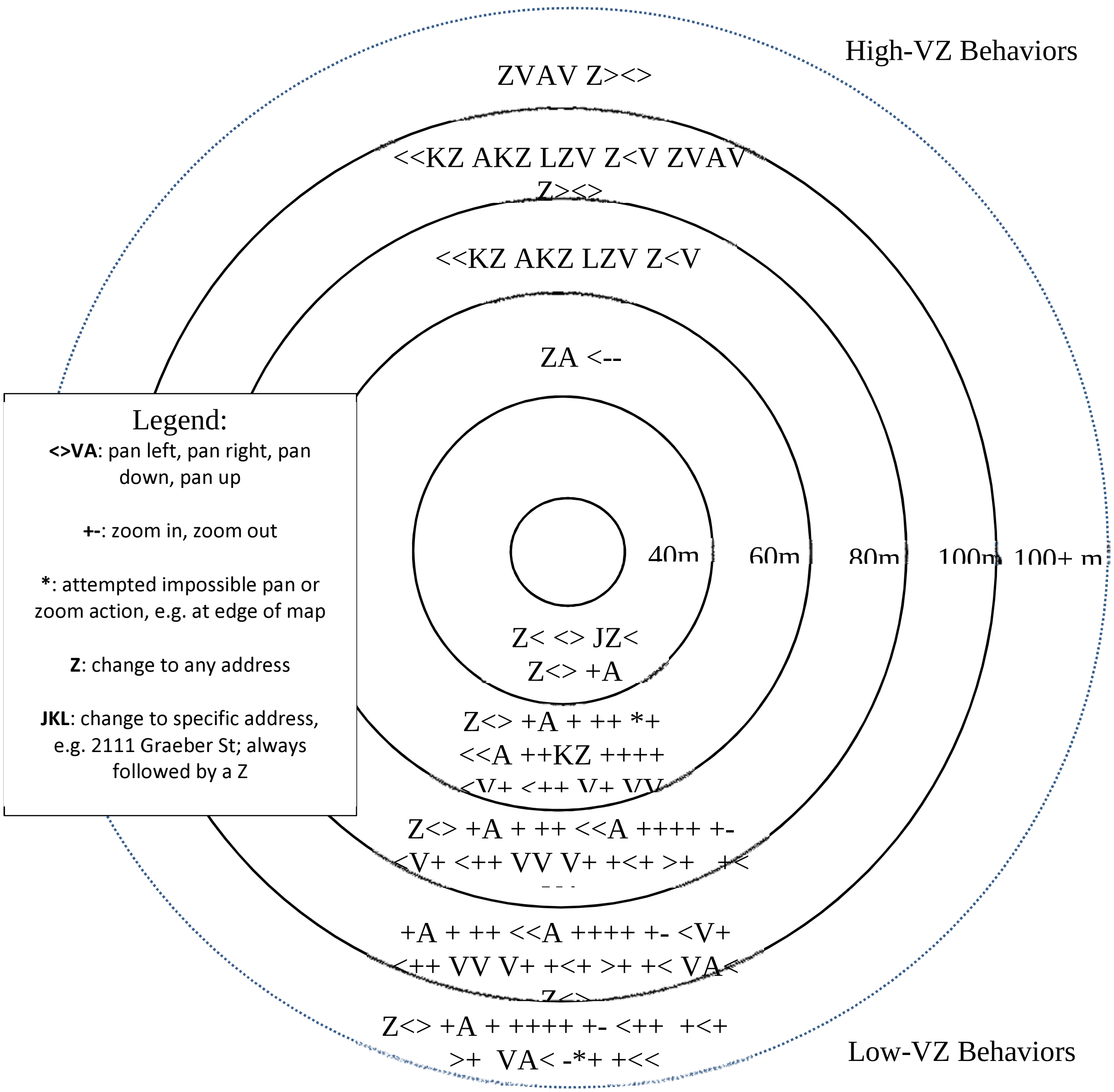


Figure 6.3 Geolocated statistically significant user behaviors in the Field and Virtual Reality experiment. The concentric circles indicate at what distance from the target address the behavior was significantly frequent. Behaviors in the top half of the graph were exhibited by high-spatial-visualization participants, while behaviors in the bottom half were exhibited by their low-spatial-visualization counterparts.

It is notable that statistically significant behavioral differences started at 25 m from target for low-spatial-visualization participants and at 50m from target for high-spatial-visualization participants. Differentials between groups manifested during the “Approach” phase. This outcome mirrors and reinforces the Paper Map Study decision model in Chapter 4, where the majority of behaviors were also observed during the “Approach” phase. Figure 6.4 contains a more fine-grained location-based presentation of behaviors.

Behavior	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100	105	Inf
ZA						x												
Z<V												x	x	x	x			
ZVAV																x	x	x
Z><>																x		
AKZ									x	x		x	x	x	x	x	x	x
<<KZ									x	x	x	x	x	x	x	x	x	x
LZV									x	x	x	x	x	x	x	x	x	x
<--							x											
Z<	x																	
<>	x	x																
JZ<	x	x																
Z<>	x	x	x	x	x	x	x	x	x	x						x	x	x
+					x	x	x	x	x	x	x	x	x	x	x	x	x	
++					x	x	x	x	x		x	x	x	x				
+A			x	x		x	x	x		x	x	x	x	x	x	x	x	
+-											x	x	x	x	x	x	x	
++++							x	x	x	x	x	x	x	x	x	x	x	x
++KZ							x											
*+					x	x	x											x
-*+																		x
V+							x	x	x	x	x	x	x	x	x		x	
<V+							x	x	x	x	x	x	x	x	x	x	x	
VV								x	x		x	x	x	x				x
<<A						x	x	x	x	x	x		x	x	x	x	x	x
VA<										x	x	x	x	x	x	x	x	x
+<						x	x			x	x	x	x	x	x	x	x	
+<+									x	x	x	x	x	x	x	x	x	x
>+										x								
+<<																		x
<++							x	x	x	x	x	x	x	x	x	x	x	x

Figure 6.4 Location-based presentation of statistically significant participant behaviors. Darker gray indicates high-spatial-visualization behaviors, while lighter gray indicates their low-spatial-visualization counterparts. Column headings show distance in meters from target address. Column heading “Inf” means “infinity”. Action symbols are explained in Table 5.2.

Another observation regarding overall differences between visualization groups is that there are fewer behaviors exhibited by high-spatial-visualization participants, again mirroring the decision model for the Paper Map Study. Furthermore, high-spatial-visualization behaviors show an emphasis on the L, K, and Z events, which are address-switching actions initiated by the user. This address-switching set of behaviors mirrors the frequent page visitation and abandonment behaviors exhibited by high-spatial-visualization participants on the search engine task in Brennan, Kelly, & Arguello (2014).

High-spatial-visualization participants also exhibited:

(a) Stepwise progression (ZVAV, Z><>): after an address submission, move the map view in a direction, move back, move forward again and remain at the new map location. This is a behavior to solidify survey knowledge while making progress;

(b) Survey information acquisition (<--). Participants were zooming out to add context.

In contrast, low-spatial-visualization participants exhibited:

(a) Magnification (+, ++, ++++, +A, +<, >+, +<+, <++ etc.): a set of behaviors that indicates a preference for a more zoomed-in map, or alternatively, for a view that minimizes survey information.

(b) Reversals (Z<>, VA<, +-, +*-): zooms or pans that were reversed and the view returned to its original location, after which a new direction might be chosen. This class of behaviors indicates searching, confusion or anchoring on a landmark.

(c) Impossible commands (*+, -*+): a set of behaviors where participants attempted to pan or zoom past the boundaries of the map and were informed that they cannot do so.

(d) Complex viewport trajectories (<<A, <V+, <++, +<+, +<<): these sequences indicate participants embarking on panning-and-zooming “expeditions” around the map, which may be the result of a zoomed-in view that minimizes survey information.

Overall, high-spatial-visualization participants exhibited survey-information-preference behaviors, while low-spatial-visualization participants exhibited survey-information-avoiding behaviors and landmark-preference behaviors, consistent with Bay and Ziefle (2008), the landmark-route-survey model of Siegel and White (1975), and the interplays between visual memory, spatial visualization ability, and working memory uncovered in Rodes & Gugerty (2012) and Meneghetti, Gyselinck, Pazzaglia, & De Beni (2009). The evidence in the literature and our data suggests the dichotomy between survey knowledge preference and landmark knowledge preference appears to map to the division between high-spatial-visualization- and low-spatial-visualization users. The next step is to recommend adaptations based on survey- and landmark-oriented publications.

6.4 Adaptation Recommendations

Buering, Gerkin and Reiterer (2006) tested high-and low-spatial-visualization participants on answering questions about movies located a scatterplot graph where each dot was a graphical object presenting the movie (Fig. 6.5). The horizontal axis denoted a popularity score, while the vertical axis denoted year of release.

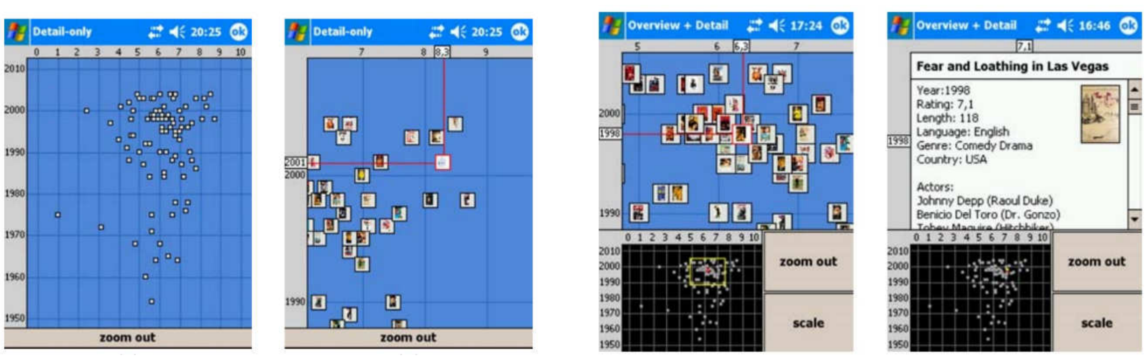


Figure 6.5 Alternative small-screen interfaces in Buering, Gerkin and Reiterer (2006). The two screens on the left do not have an overview window, while the two screens on the right are from the interface with an overview window. High-spatial-visualization participants performed better without the overview window, while low-spatial-visualization participants performed better with the overview window.

The experimental adaptation was an **overview window** that shared screen real estate with the freely zoomable graph view (Figure 6.5, right). High-spatial-visualization participants took longer to complete the experiment with an overview window (as compared to having no overview window), while low-spatial-visualization participants took longer to complete the experiment without an **overview window** (as compared to having an overview window).

Delikostidis, Elzakker, & Kraak (2016)'s study elicited user requirements for two mobile location-based applications. The authors reported that landmark visuals "helpful for orientation and navigation were particularly road patterns and sizes, street names, parks/squares and roundabouts. Helpful but not always visible landmarks on the map were: bridges, pedestrian paths, and important or tall buildings visible from a distance". Additionally, the authors clarified that "Popped-up photos of landmarks were regarded as more helpful than their 3D representations". The authors' observations can be incorporated into an interface adaptation with **enlarged landmarks** in the map view to draw the attention of the user. Enlargement may be based on proximity to the user. The **enlarged landmarks** should be accompanied by muted visual presentation of the rest of the map in order to avoid overwhelming the user's spatial visualization capability.

Stanney, Chen, Wedell, and Breaux (2003, pp. 213-214) propose a visualization of **timestamped waypoints** to aid in recovering orientation. An adaptation that targets low-spatial-visualization users could go one step further and continually display the entire route from the start of the work session, with **timestamped and connected visited landmarks**.

Willis, Hölscher, Wilbertz, & Li (2009) tested participants on their survey knowledge of an environment after having explored it with a paper map or a mobile phone map. Mobile phone participants took 46 minutes on average to familiarize themselves with the environment while walking inside it on a predetermined path. In contrast, paper map users studied a map for an average of 18 minutes and never set foot in the actual space. At the end of training all participants were taken to a location within the environment and asked to provide direction and distance estimates to various targets. Mobile phone participants were unable to perform at the level of their paper map counterparts despite taking much longer to familiarize themselves with the environment. One problem discovered by the authors was that mobile map users had a relatively passive interaction with the software map due to having to follow a predetermined route. The authors suggested that participants **confirm information about the area** while traversing the route, in order to keep alert and remain engaged in learning the spatial layout. Conversely, extra survey-level confirmations required of low-spatial-visualization users might overwhelm and bewilder them. Landmark-level confirmations may provide a benefit instead. This adaptation also addresses observed behaviors in the Field and Virtual Reality Study where high-spatial-visualization participants overlooked important environmental cues and solved tasks incorrectly due to an overly hasty approach to the experiment.

Willis, Hölscher, Wilbertz, & Li (2009) also identified unstable cognitive schemata resulting from fragmented survey knowledge acquired from a small screen. The suggested remedy was to enable **pre-planning on suitably zoomed out representations**, which should be **revisited** periodically to solidify the connection between fragments. This adaptation aligns with planning behaviors observed in the Paper Map study, as well as with a high-VZ software behavior from the Field and Virtual Reality study: “pan-left, zoom-out, zoom-out”. In contrast, low-VZ software behaviors from the same study were overwhelmingly composed of zoom-in actions. Therefore, a planning adaptation could inconvenience low-spatial-visualization participants considerably if their preference is to absorb information in smaller chunks.

6.5 Conclusion

Reports in the literature and data from the Field and Virtual Reality Study suggested that high-spatial-ability users would be best served by adaptations enhancing the availability of survey information. Conversely, low-spatial-visualization participants should benefit from landmark-oriented adaptations. More adaptation experiments targeting spatial visualization ability are needed to fill the gap in current understanding.

CHAPTER 7. CONCLUSION

In this dissertation, we described a novel, behavior-based spatial visualization detection technique that can be adopted on generic computer devices.

Other user-sensing results in the current literature exhibit two or more of the following conditions: (a) investigating volatile (rather than semi-permanent) cognitive variables such as learning and frustration; (b) relying on one or more sensors such as galvanic skin response, pressure, and gaze tracking; (c) detection rates in the 70% range and no higher; (d) user variables not connected with personal computer ability; and (e), results not related to classification of future users.

In contrast, our results allow the detection of a semi-permanent cognitive ability that is known to relate to users' computer performance. The detection technique does not depend on sensors and is immediately deployable on generic computers. Detection rates exceeded 80% and approached 90%. The algorithm does not need to know whether the user made mistakes, so detection can be unaware of environmental conditions, and therefore need no sensors. User ability is detectable on a complex, location-based, professional task without any knowledge of the environment, thereby reducing both hardware and software costs through simplification.

The detection of spatial visualization ability allows coupling with ability-specific software adaptations. Sources in the literature indicate that low-spatial-visualization users benefit from landmark-oriented adaptations, while high-spatial-visualization users prefer survey-oriented adaptations. Experimental data and published reports imply user-selected adaptations cannot be guaranteed to enhance performance due to the subconscious nature of individual differences. An automatic solution should be implemented instead.

The present study is an early step towards operationalizing the close relationship between spatial visualization ability and users' computer activities. An immediate next step is to build an adaptive system and validate the current design expectations. More participants are necessary! While the literature does not present any detraction to the proposed adaptations, they will strongly influence the workflow and must be certified.

More rigorous feature engineering could be applied to the detection technique in order to bring detection accuracy higher. The available data includes precise timing and cursor movement traces that are yet untapped.

Medium-term goals include determining a more detailed composition of the user base with regard to spatial visualization. In particular, there are indicators of several sub-groups with varying levels of ability and behavior. A multi-participant user study is necessary to bring out sufficient representation of all cohorts along the full range of spatial visualization. The project is expected to be complicated by interference from other user characteristics.

The present detection technique need not be limited to location-based interfaces. While map-centered systems all but guarantee spatial visualization ability is pertinent, the literature is unequivocal with regard to the ability playing a role in multiple other task types such as advanced information retrieval and remote teleoperation of robots.

In the long term, new developments in the psychology of spatial knowledge, e.g. Meneghetti, Labate, Pazzaglia, Hamilton, & Gyselinck (2016), invite embracing a more complex model of visuospatial processing, with multiple individual differences, and subsequently discovering the boundaries of the relevant design space. The outcome of such extensive activity would be cognition-aware software engineering, and cognitively-tuned interfaces.

REFERENCES

About Us, what the U.S. Census Bureau does. http://www.census.gov/about/what.html. Accessed 1/17/2015.

Ackerman, P. Ability Determinants of Individual Differences in Skilled Performance. In R. J. Sternberg and J.E. Pretz (eds). Cognition and Intelligence: Identifying the Mechanisms of the Mind. Cambridge University Press. Cambridge, UK. 2005, pp. 142-159

Alonso, D. L., Rose, A., Plaisant, C., & Norman, K. L. (1998). Viewing personal history records: A comparison of tabular format and graphical presentation using LifeLines. *Behaviour & Information Technology, 17*(5), 249-262.

Andrews, E. C., & Westerman, S. J. (2012). Age differences in simulated driving performance: Compensatory processes. Accident Analysis & Prevention, 45(0), 660–668.

Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. Educational Technology Research and Development, 1–16.

Argenta, C., & Hale, C. R. (2015). Analyzing Variation of Adaptive Game-Based Training with Event Sequence Alignment and Clustering. In *Proceedings of the Third Annual Conference on Advances in Cognitive Systems Poster Collection* (Article 26).

Batinov, G., Whitney, K., Miller, L., Nusser, S., Stanfill, B., & Ashenfelter, K. T. (2013). Evaluating the Impact of Spatial Ability in Virtual and Real World Environments. Sixth International Conference on Advances in Computer-Human Interactions (ACHI 2013). Nice, France. February 24 - March 1, 2013.

Bay, S., & Ziefle, M. (2008). Landmarks or surveys? The impact of different instructions on children's performance in hierarchical menu structures. *Computers in Human Behavior*, *24*(3), 1246-1274.

Benyon, D., Crerar, A., & Wilkinson, S. (2001). Individual differences and inclusive design. In C. Stephanidis (Ed'.), User interfaces for all: concepts, methods, and tools. Mahwah, NJ: Lawrence Erlbaum Associates.

Bixler, R., & D'Mello, S. (2015). Automatic gaze-based user-independent detection of mind wandering during computerized reading. *User Modeling and User-Adapted Interaction*, 1-36.

Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial intelligence*, *97*(1), 245-271.

Bohnenberger, T., Jameson, A., Kruger, A., & Butz, A. (2002). Location-Aware Shopping Assistance: Evaluation of a Decision-Theoretic Approach. Lecture Notes in Computer Science, 151-169.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*(2), 123-140.

Brennan, K., Kelly, D., & Arguello, J. (2014). The effect of cognitive abilities on information search for tasks of varying levels of complexity. In *Proceedings of the 5th Information Interaction in Context Symposium* (pp. 165-174). ACM.

Buering, T., Gerken, J., & Reiterer, H. (2006). User interaction with scatterplots on small screens-a comparative evaluation of geometric-semantic zoom and fisheye distortion. *IEEE Transactions on Visualization and Computer Graphics*, *12*(5), 829-836.

Bunt, A., Terry, M., & Lank, E. (2013). Challenges and Opportunities for Mathematics Software in Expert Problem Solving. Human–Computer Interaction, 28(3), 222-264.

Calvet Liñán, L., & Juan Pérez, Á. A. (2015). Educational Data Mining and Learning Analytics: differences, similarities, and time evolution. RUSC. Universities and Knowledge Society Journal, 12(3). pp. 98-112. doi: http://dx.doi.org/10.7238/rusc.v12i3.2515

Calvo, R. A., & D'Mello, S. (2010). Affect Detection: An Interdisciplinary Review of Models, Methods and their Applications. IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, 1(1).

Campagni, R., Merlini, D., Sprugnoli, R. & Verri, M. C. (2015). Data mining models for student careers. *Expert Systems With Applications*, *42*(13), 5508-5521.

Campagnoni, F. R., & Ehrlich K. (1989). Information retrieval using a hypertext-based help system. Transactions on Information Systems, 7, 271–291.

Chang, J. M., Fang, C. C., Ho, K. H., Kelly, N., Wu, P. Y., Ding, Y., Chu, C.,Gilbert, S., Kamal, A.E., & Kung, S. Y. (2013). Capturing cognitive fingerprints from keystroke dynamics. IT Professional, 15(4), 24-28.

de Campos, L. M., Fernández-Luna, J. M., Huete, J. F., & Vicente-Lopez, E. (2014). Using personalization to improve XML retrieval. *Knowledge and Data Engineering, IEEE Transactions on*, *26*(5), 1280-1292.

del Puerto Paule-Ruiz, M., Riestra-Gonzalez, M., Sánchez-Santillan, M., & Pérez-Pérez, J. R. (2015). The Procrastination Related Indicators in e-Learning Platforms. *Journal of Universal Computer Science*, *21*(1), 7-22.

Delikostidis, I., van Elzakker, C. P., & Kraak, M. J. (2016). Overcoming challenges in developing more usable pedestrian navigation systems. *Cartography and Geographic Information Science*, *43*(3), 189-207.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, *55*(10), 78-87.

Downing, R. E., Moore, J. L., & Brown, S. W. (2005). The effects and interaction of spatial visualization and domain expertise on information seeking. Computers in Human Behavior, 21(2), 195–209.

Dumas, B., Lalanne, D., & Oviatt, S. (2009). Multimodal Interfaces: A Survey of Principles, Models and Frameworks. In D. Lalanne, & J. Kohlas (Eds.), Human Machine Interaction, LNCS 5440 (pp. 3-26). Berlin Heidelberg: Springer-Verlag.

Ekstrom, R. B., French, J. W., Harman, H. H. , & Dermen, D. (1976). Manual for kit of factor-referenced cognitive tests. Princeton, NJ: Educational Testing Service.

Ezen-Can, A., Grafsgaard, J. F., Lester, J. C., & Boyer, K. E. (2015). Classifying student dialogue acts with multimodal learning analytics. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (pp. 280-289). ACM.

Frank, E., Wang, Y., Inglis, S., Holmes, G., & Witten, I.H. (1998)."Using model trees for classification." *Machine Learning* 32.1. 63-76.

Froese, M. E., Tory, M., Evans, G. W., & Shrikhande, K. (2013). Evaluation of static and dynamic visualization training approaches for users with different spatial abilities. *IEEE transactions on visualization and computer graphics*, *19*(12), 2810-2817.

Galán, F. C., & Beal, C. R. (2012). EEG Estimates of Engagement and Cognitive Workload Predict Math Problem Solving Outcomes. User Modeling, Adaptation, and Personalization, Lecture Notes in Computer Science,Vol. 7379, pp. 51–62.

Gleasure, R., Feller, J., & O'Flaherty, B. (2012). Human-centered design: existing approaches and a future research agenda. ECIS 2012 Proceedings (p. Paper 105). http://aisel.aisnet.org/ecis2012/105.

Gonzalez-Rodriguez, M., Manrubia, J., Vidau, A., & Gonzalez-Gallego, M. (2009). Improving accessibility with user-tailored interfaces. Appl Intell, 30(1), 65-71.

Hau Chua, S., Zhang, H., Hammad, M., Zhao, S., Goyal, S., & Singh, K. (2015). ColorBless: Augmenting visual information for colorblind people with binocular luster effect. ACM Transactions on Computer-Human Interaction (TOCHI), 21(6), 32.

Hong, J.-y., Suh, E.-h., & Kim, S.-j. (2009). Context-aware systems: A literature review and classification. Expert Systems with Applications, 36(4), 8509–8522.

Höök, K. (1997). Evaluating the utility and usability of an adaptive hypermedia system. Knowledge-Based Systems, 10(5), 311-19.

Höök, K. (2000). Steps to take before intelligent user interfaces become real. Interacting with Computers, 12, 409-426.

Hourcade, J., Nguyen, C. M., Perry, K. B., & Denburg, N. L. (2010). PointAssist for Older Adults: Analyzing Sub-Movement Characteristics to Aid in Pointing Tasks. Computer-Human Interaction. Atlanta, GA, USA: ACM.

Jennings, F., Benyon, D. R., & Murray, D. M. (1991). Adapting systems to differences between individuals. Acta Psychologica, 78(1-3), 243–258.

Kinnebrew, J. S., & Biswas, G. (2012). Identifying Learning Behaviors by Contextualizing Differential Sequence Mining with Action Features and Performance Evolution. Proceedings of the 5th International Conference on Educational Data Mining., 57–64.

Kozhevnikov, M., Motes, M. A., Rasch, B., & Blajenkova, O. (2006). Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance. Applied Cognitive Psychology, 20, 397-417.

Knutov, E., De Bra, P., & Pechenizkiy, M. (2009). AH 12 years later: a comprehensive survey of adaptive hypermedia methods and techniques. New Review of Hypermedia and Multimedia, 15(1), 5-38.

Koedinger, K. R., McLaughlin, E. A., & Stamper, J. C. (2012). Automated Student Model Improvement. Proceedings of the 5th International Conference on Educational Data Mining. pp. 17–24.

Kotzyba, M., Siegert, I., Gossen, T., Wendemuth, A., & Nürnberger, A. (2015). Exploratory Voice-Controlled Search for Young Users: Challenges & Potential Benefits. http://www.researchgate.net/profile/Andreas_Nuernberger/publication/274838463_Exploratory_Voice-Controlled_Search_for_Young_Users__Challenges_and_Potential_Benefits/links/55312b1d0cf2f2a588ad1a36.pdf. Accessed 11/5/2015.

Krogh, A., & Vedelsby, J. (1995). Neural network ensembles, cross validation, and active learning. *Advances in neural information processing systems*, *7*, 231-238.

Lee, D. Y., & Shin, D. H. (2011). Effects of spatial ability and richness of motion cue on learning in mechanically complex domain. *Computers in Human Behavior*, *27*(5), 1665-1674.

Liu, A. M., Oman, C. M., Galvan, R., & Natapoff, A. (2012). Predicting space telerobotic operator training performance from human spatial ability assessment. Acta Astronautica, (0), -. doi:10.1016/j.actaastro.2012.04.004

Manyika, Chui, Brown, Bughn, Dobbs, Roxburgh, Byers. (2012). Big data: The next frontier for innovation, competition, and productivity. McKinsey Institute.

Maxion, R. A., & Reeder, R. W. (2005). Improving user-interface dependability through mitigation of human error, International Journal of Human-Computer Studies, Volume 63, Issues 1–2, July 2005, Pages 25-50, ISSN 1071-5819, http://dx.doi.org/10.1016/j.ijhcs.2005.04.009.

Meneghetti, C., Gyselinck, V., Pazzaglia, F., & De Beni, R. (2009). Individual differences in spatial text processing: High spatial ability can compensate for spatial working memory interference. *Learning and Individual differences*, *19*(4), 577-589

Meneghetti, C., Labate, E., Pazzaglia, F., Hamilton, C., & Gyselinck, V. (2016). The role of visual and spatial working memory in forming mental models derived from survey and route descriptions. *British Journal of Psychology*.

Mirza, H. T., Chen, L., Hussain, I., Majid, A., & Chen, G. (2015). A Study on Automatic Classification of Users' Desktop Interactions. Cybernetics and Systems, 46(5), 320-341.

Modjeska, D., & Chignell, M. (2003). Individual differences in exploration using desktop VR. *Journal of the American Society for Information Science and Technology*, *54*(3), 216-228.

Nguyen, N. T. (2012). *Anatomy: The Relationship Between Internal and External Visualizations* (Doctoral dissertation, The University of Western Ontario).

Norman, D. A., & Draper, S. W. (1986). User Centered System Design: New Perspectives on Human-Computer Interaction. Hillsdale, NJ, USA: L. Erlbaum Associates Inc.

Ohm, C., Bienk, S., Kattenbeck, M., Ludwig, B., & Müller, M. (2016). Towards interfaces of mobile pedestrian navigation systems adapted to the user's orientation skills. *Pervasive and Mobile Computing*, *26*, 121-134.

Parasuraman, R. (2003). Neuroergonomics: Research and practice. *Theoretical Issues in Ergonomics Science*, *4*(1-2), 5-20.

Peterson, J., Pardos, Z., Rau, M., Swigart, A., Gerber, C., & McKinsey, J. (2015). Understanding Student Success in Chemistry using Gaze Tracking & Pupillometry. In Lecture Notes in Computer Science, Vol. 9112, pp. 358-366. Springer Berlin / Heidelberg.

Quinlan, John R. (1992)."Learning with continuous classes." *5th Australian joint conference on artificial intelligence*. Vol. 92. pp.343-348.

Rath, A. S. (2010). User Interaction Context: Studying and Enhancing Automatic User Task Detection on the Computer. Dissertation. Graz University of Technology.

Rath, A., Devaurs, D., & Lindstaedt, S. (2010). Studying the Factors Influencing Automatic User Task Detection on the Computer Desktop. In M. Wolpers, P. Kirschner, M. Scheffel, S. Lindstaedt, & V. Dimitrova (Eds.), Lecture Notes in Computer Science (Vol. 6383, pp. 292–307). Springer Berlin / Heidelberg. Retrieved from http://dx.doi.org/10.1007/978-3-642-16020-2_20

Rodes, W., & Gugerty, L. (2012). Effects of Electronic Map Displays and Individual Differences in Ability on Navigation Performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*(4), 589–599.

Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews, 40(6), 601-618.

Rothrock, L., Koubek, R., Fuchs, F., Haas, M., & Salvendy, G. (2002). Review and reappraisal of adaptive interfaces: toward biologically inspired paradigms. Theoretical Issues in Ergonomics Science, 3(1), 47-84.

Rusch, M. L., Nusser, S. M., Miller, L. L., Batinov, G. I, & Whitney, K. (2012). Evaluating the Impact of Spatial Ability in Virtual and Real World Environments. Proceedings of the Fifth International Conference on Advances in Computer-Human Interactions (ACHI 2012). pp. 35-40.

Schaffer, S., Schleicher, R., & Möller, S. (2015). Modeling input modality choice in mobile graphical and speech interfaces. *International Journal of Human-Computer Studies*, *75*, 21-34.

Sciarini, W. L., Fidopiastis, C., & Nicholson, D. (2009, October). Towards a modular cognitive state gauge: Assessing spatial ability utilization with multiple physiological measures. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, No. 3, pp. 146-150). SAGE Publications.

Sciarini, L. W., Grubb, J. D., & Fatolitis, P. G. (2014, September). Cognitive State Assessment Examination of EEG-based Measures on a Stroop Task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 215-219). SAGE Publications.

Seagull, F. & Walker, N. (1992). The effects of hierarchy structure and visualization ability on computerized information retrieval. International Journal of Human Computer Interaction, 4, 369– 385.

Siegel, A. W., & White, S. H. (1975). The development of spatial representations of large-scale environments. *Advances in child development and behavior*, *10*, 9-55.

Stanney, K. M., Chen, J. L., Wedell, B., & Breaux, R. (2003). Identification of metaphors for virtual environment training systems. *Ergonomics*, *46*(1-3), 197-219.

Steichen, B., Ashman, H., & Wade, V. (2012). A comparative survey of Personalised Information Retrieval and Adaptive Hypermedia techniques. Information Processing Management, 48(4), 698 – 724. doi:10.1016/j.ipm.2011.12.004

Taylor, Sr., A., Miller, L., Nilakanta, S., Sander, J., Mitra, S., Sharda, A., et al. (2009). Using an Error Detection Strategy for Improving Web Accessibility for Older Adults. Advances in Computer-Human Interactions, (pp. 375 - 380). Cancun.

Tewes, U. (1983). Hamburg–Wechsler Intelligenztest fu¨r Kinder [Hamburg–Wechsler intelligence test]. Revised. (HAWIK-R). Bern, Switzerland: Huber.

Thomas, P., Bailey, P., Moffat, A., & Scholer, F. (2015). Towards Nuanced System Evaluation Based on Implicit User Expectations. Proceedings of the 11th Asia Information Retrieval Societies Conference, 332-344.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and motor skills*, *47*(2), 599-604.

Van Velsen, L., Van Der Geest, T., Klaassen, R., & Steehouder, M. (2008). User-centered evaluation of adaptive and adaptable systems: a literature review. The Knowledge Engineering Review, 23(3), 261-281.

Vicente, K. J., Hayes, B. C., & Williges, R. C. (1987). Assaying and isolating individual differences in searching a hierarchical file system. Human Factors: The Journal of the Human Factors and Ergonomics Society, 29(3), 349–359.

Waalkens, M., Aleven, V., & Taatgen, N. (2013). Does supporting multiple student strategies lead to greater learning and motivation? Investigating a source of complexity in the architecture of intelligent tutoring systems. Computers & Education, 60(1), 159–171. doi:10.1016/j.compedu.2012.07.016

Wang, Y., & Witten, I.H. (1997). "Inducing Model Trees for Continuous Classes." *In Proc. of the 9th European Conf. on Machine Learning Poster Papers*. Prague, Czechoslovakia.

Whitney, K. C., Batinov, G. I., Nusser, S. M., Miller, L. L., & Ashenfelter, K. T. (2011). Exploring a map survey task's sensitivity to cognitive ability. Paper presented at the The Fourth International Conference on Advances in Computer-Human Interactions, Gosier, Guadeloupe, France.

Willis, K. S., Hölscher, C., Wilbertz, G., & Li, C. (2009). A comparison of spatial knowledge acquisition with maps and mobile maps. *Computers, Environment and Urban Systems*, *33*(2), 100-110.

Williges, R. C., Elkerton, J., Vicente, K. J., & Hayes, B. C. (1990). *Adaptive Human-Computer Interfaces Using Expert Profiles* (No. HCI-87-01). VIRGINIA POLYTECHNIC INST AND STATE UNIV BLACKSBURG.

Wilson, M. W. (2012). Location-based services, conspicuous mobility, and the location-aware future, Geoforum, Volume 43, Issue 6, November 2012, Pages 1266-1275, ISSN 0016-7185, 10.1016/j.geoforum.2012.03.014.

Wolpert, D. H. (1992). Stacked generalization. *Neural networks*, *5*(2), 241-259.

Worsley, M., & Blikstein, P. (2015). Leveraging multimodal learning analytics to differentiate student learning strategies. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (pp. 360-367). ACM.

Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior, 47*, 168-181.

Yen, G. G., & Acay, D. (2009). Adaptive user interfaces in complex supervisory tasks. ISA Transactions, 196-205.

Yoon, S. Y., Choi, Y. J., & Oh, H. (2015). User attributes in processing 3D VR-enabled showroom: Gender, visual cognitive styles, and the sense of presence. *International Journal of Human-Computer Studies*, *82*, 1-10.

Zhang, H., & Salvendy, G. (2001). The implications of visualization ability and structure preview design for Web information search tasks. International Journal of Human-Computer Interaction, 13(1), 75–95.

APPENDIX A. PAPER MAP STUDY: SCHEDULING SCRIPTS AND ADVERTISING

(continued on next page)

ISU IRB # 1	09-386
Approved Date:	28 September 2009
Expiration Date:	12 September 2010

PARTICIPANTS NEEDED

to verify street addresses in a neighborhood for research study.

Participants must be ISU students, aged 18 or older, and fluent in English. Participants should have minimal exposure to Ames neighborhoods.

Participants will be offered compensation.

For more information or to schedule an appointment, please contact:

KOFI WHITNEY
Email: kwhitney@iastate.edu
Phone: (803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

Address Verification Study
Kofi: kwhitney@iastate.edu
(803) 546-0007

ISU IRB # 1 09-386
Approved Date: 28 September 2009
Expiration Date: 12 September 2010

Newspaper advertisement

Participants Needed for Research Study

We are looking for participants to verify street addresses in a neighborhood. Participants must be ISU students, aged 18 or older, and fluent in English. Participants should have minimal exposure to Ames neighborhoods.

Participants completing the study will be offered compensation.

For more information or to schedule an appointment, please contact:
Kofi Whitney @ kwhitney@iastate.edu or (803) 546-0007.

ISU IRB # 1	09-386
Approved Date:	28 September 2009
Expiration Date:	12 September 2010

SCHEDULING SCRIPTS (continued)

Field Exercise – Email

Subject: Address Verification Study - Field Exercise

Dear [*Student's Name*],

You have received this email because you participated in the screening portion of our Address Verification Study. We have reviewed your screening information and would like to invite you to participate in the field exercise portion of the study. This exercise will take approximately 2 hours to complete. You will receive a $20 gift card for your participation.

Please refer to your *Informed Consent* document for additional information regarding the study. You are welcome to contact us if you would like to receive another copy.

Click here [*Doodle Scheduling Link*] to schedule the field exercise.
Note: We may contact you to reschedule if weather conditions are not favorable.

Thank you on behalf of my research group for participating in this study.

Sincerely,

Kofi Whitney
Graduate Assistant
Department of Computer Science

Your participation in this study is completely voluntary and you may withdraw at any time. The data that we collect from you will be kept confidential. If you have any questions or concerns about this study, please contact Kofi Whitney @ 803-546-0007 or Dr. Les Miller @ 515-294-7934.

ISU IRB # 1	09-366
Approved Date:	25 September 2009
Expiration Date:	12 September 2010

SCHEDULING SCRIPTS

The Doodle Online scheduling tool will be used – http://www.doodle.com
After a student has been scheduled, a follow up email will be sent with appointment information

Screening – Email

Subject: Address Verification Study - Screening

Dear [*Student's Name*],

You have received this email because you have expressed an interest to participate in our Address Verification Study. The next step is to schedule you for the screening portion of the study. The screening will involve a background information questionnaire and a series of cognitive tests. Screening will take approximately 1 hour to complete. You will receive a $10 gift card for participating in the screening.

You may be selected after this screening to later participate in a field exercise that will take approximately 2 hours to complete. We ask that you schedule this initial screening if and only if you intend to participate in the field exercise. Compensation for the field exercise is a $20 gift card.

Click here [*Doodle Scheduling Link*] to schedule a screening appointment.

Thank you on behalf of my research group for participating in this study.

Sincerely,

Kofi Whitney
Graduate Assistant
Department of Computer Science

Your participation in this study is completely voluntary and you may withdraw at any time. The data that we collect from you will be kept confidential. If you have any questions or concerns about this study, please contact Kofi Whitney @ 803-546-0007 or Dr. Les Miller @ 515-294-7934.

Screening – Phone

- "Hello, my name is [Scheduler's Name]. I am calling you because you have expressed an interest to participate in our Address Verification Study. I'd like to remind you that your participation in this study is completely voluntary and you may withdraw at any time. The data that we collect from you will be kept confidential."
- "May I continue?"
- "The next step is to schedule you for the screening portion of the study. The screening will involve a background information questionnaire and a series of cognitive tests. Screening will take approximately 1hour to complete. You will receive a $10 gift card for participating in the screening. You may be selected after this screening to later participate in a field exercise that will take approximately 2 hours to complete. We ask that you schedule this initial screening if and only if you intend to participate in the field exercise. Compensation for the field exercise is a $20 gift card."
- "May I email you the information that I have discussed along with a link that will allow you to schedule your screening appointment?"
- "Thank you for your time."

APPENDIX B. PAPER MAP STUDY: INFORMED CONSENT FORM

(continued on next page)

ISU IRB # 1	09-386
Approved Date:	28 September 2009
Expiration Date:	12 September 2010

INFORMED CONSENT

The purpose of this research study is to gather information on how individuals compare addresses on a street with information on a paper map.

You will complete a brief background questionnaire followed by 3 cognitive assessments. This will take approximately 1 hour. Some subjects will be contacted at a later date to schedule an appointment for a field exercise. If selected for the field exercise, we will explain the task and train you on the procedures. You will then be transported, via CyRide, to an Ames neighborhood where you will practice the procedures. Next, we will give you a list of addresses to find and verify against a map that we provide. During the field exercise, you will be asked to think aloud as you reason through the task and this will be audio-recorded. This exercise will take approximately 2 hours to complete. There are no known risks to participation other than concerns that are normally associated with walking through a neighborhood.

There are no direct benefits to you as a participant other than the educational experience of being involved in an study. Your participation is helping us learn how we can improve address listing methods. By participating in this study, you will be offered a $10 gift card for the background questionnaire and cognitive assessments. If you are selected for the field exercise, you will be offered an additional $20 gift card for participation. You will need to sign a receipt for both gift cards.

Your participation in this study is completely voluntary and you may withdraw at any time. You may skip any part of this study that makes you feel uncomfortable or withdraw from the study at any time without penalty or loss of benefits to which you may otherwise be entitled.

Records identifying participants will be kept confidential to the extent permitted by applicable laws and regulations and will not be made publicly available. However, federal government agencies, the National Science Foundation, and the Institutional Review Board (a committee that reviews and approves human subject research studies) may inspect and/or copy your records for quality assurance and data analysis. These records may contain private information. To ensure confidentiality to the extent permitted by law, the following measures will be taken:

Data that identifies participants will be kept confidential. The information taken from this session will be assigned a unique code. Your name will not be associated with this information. Only researchers from Iowa State University working on this project will have access to data collected during this study. Study records will be kept confidential under password protected computer files. Data will be retained for two years and then will be destroyed.

The data results from this research may be used for educational or scientific purposes and may be presented at scientific and/or educational meetings or published in professional journals. Results will be released in summary form only with no personal identifying information.

For further information about the study contact Kofi Whitney at (803) 546-0007 or Dr. Les Miller at (515) 294-7588. If you have any questions about the rights of research subjects or research-related injury, please contact the IRB Administrator, (515) 294-4566, IRB@iastate.edu, or Director, Office for Responsible Research, (515) 294-3115, 1138 Pearson Hall, Ames, IA 50011.

Your signature indicates that you voluntarily agree to participate in this study, that the study has been explained to you, that you have been given the time to read the document and that your questions have been satisfactorily answered. You will receive a copy of the written informed consent.

Participant's Name (printed)____

______________________________ ______________________
(Participant's Signature) (Date)

INVESTIGATOR STATEMENT

I certify that the participant has been given adequate time to read and learn about the study and all of their questions have been answered. It is my opinion that the participant understands the purpose, risks, benefits and the procedures that will be followed in this study and has voluntarily agreed to participate.

______________________________ ______________________
(Signature of Person Obtaining Informed Consent) (Date)

APPENDIX C. PAPER MAP STUDY: TRAINING SCRIPT

(continued on next page)

ISU IRB # 1 09-386
Approved Date: 28 September 2009
Expiration Date: 12 September 2010

TRAINING SCRIPT

Participants will read along as this is dictated by the facilitator

You will be presented with the following training materials: (1) a map of the residential area and (2) a list of addresses.

You will be asked to determine whether the addresses are accurately reflected on the map. You may verify the list addresses in any order that you prefer. Four outcomes are possible during verification: (1) the ground situation is correctly reflected on the map; (2) the map erroneously displays a housing unit that is not on the ground; (3) the map erroneously displays a housing unit that is on the ground but incorrect; (4) the map does not display a housing unit appearing on the ground. Procedures for modifying your map will now be outlined by the facilitator.

A "think-aloud" method will be used during this exercise. You have been equipped with an audio recording device. You will be asked to verbalize your thoughts about performing the task throughout the exercise. These thoughts will be recorded to help us accurately recall your approach.

We ask that you perform this exercise using your typical practices for interpreting a map and identifying residential homes. It is important that you say aloud everything that you think or do. Do not feel uncomfortable or embarrassed about your approach or any of your thoughts. Even the minute pieces of information that you provide are important to us and all of your input will be held in the strictest of confidence.

Your facilitator's role will be only to observe and record your behavior. They will not interact with you other than to encourage you to think aloud or to get clarification on something that you said that cannot be verbally understood. We will begin by a mock exercise to acclimate you to the think-aloud process.

Once you have completed the training, you will then move on to the main exercise. It will be conducted in the same manner as the training—you should expect your observer to be completely passive at this point unless you have stopped verbalizing your thoughts. Before beginning the main exercise, please ensure that you are comfortable with the procedure and that all of your questions have been answered.

APPENDIX D. PAPER MAP STUDY: OBSERVER CODING SHEET

(continued on next page)

ISU IRB # 1 09-386
Approved Date: 28 September 2009
Expiration Date: 12 September 2010

CODING SHEET

Date: ________ Observer: ________ ID:

Address:____ ____ Order:____ Action: **V A M D** Incidental verification ❑ Time taken:_____

Map

Check: ______________________

Rotation: ______________________

Modification: ______________________

Body

Heading: ______________________

____

Body rotation: ______________________

Street

Sign check: ______________________

Numbering: ______________________

Odd/even: ______________________

Other: ______________________

____

Miscellaneous

Confusion sources: ______________________

____

List modification: ______________________

____

NOTES

APPENDIX E. EKSTROM ET AL. (1976) TEST OF SPATIAL VISUALIZATION

(continued on next page)

Paper Folding Test -- VZ-2

Suggested by Thurstone's Punched Holes. For each item successive drawings illustrate two or three folds made in a square sheet of paper. The final drawing of the folded paper shows where a hole is punched in it. The subject selects one of 5 drawings to show how the punched sheet would appear when fully reopened.

Length of each part: 10 items, 3 minutes

Suitable for grades 9-16

ISU IRB # 1 09-386
Approved Date: 28 September 2009
Expiration Date: 12 September 2010

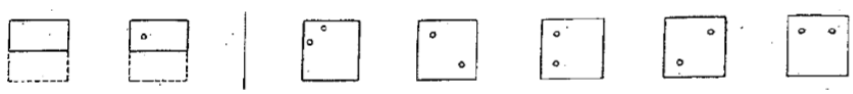
Name ______________________________

PAPER FOLDING TEST — VZ-2

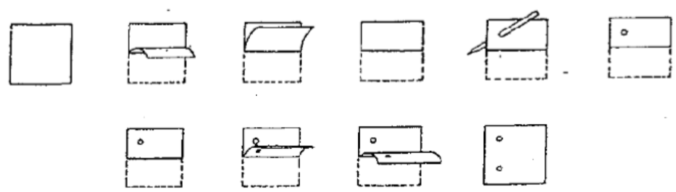
In this test you are to imagine the folding and unfolding of pieces of paper. In each problem in the test there are some figures drawn at the left of a vertical line and there are others drawn at the right of the line. The figures at the left represent a square piece of paper being folded, and the last of these figures has one or two small circles drawn on it to show where the paper has been punched. Each hole is punched through all the thicknesses of paper at that point. One of the five figures at the right of the vertical line shows where the holes will be when the paper is completely unfolded. You are to decide which one of these figures is correct and draw an X through that figure.

Now try the sample problem below. (In this problem only one hole was punched in the folded paper.)

A B C D E



The correct answer to the sample problem above is C and so it should have been marked with an X. The figures below show how the paper was folded and why C is the correct answer.



In these problems all of the folds that are made are shown in the figures at the left of the line, and the paper is not turned or moved in any way except to make the folds shown in the figures. Remember, the answer is the figure that shows the positions of the holes when the paper is completely unfolded.

Your score on this test will be the number marked correctly minus a fraction of the number marked incorrectly. Therefore, it will <u>not</u> be to your advantage to guess unless you are able to eliminate one or more of the answer choices as wrong.

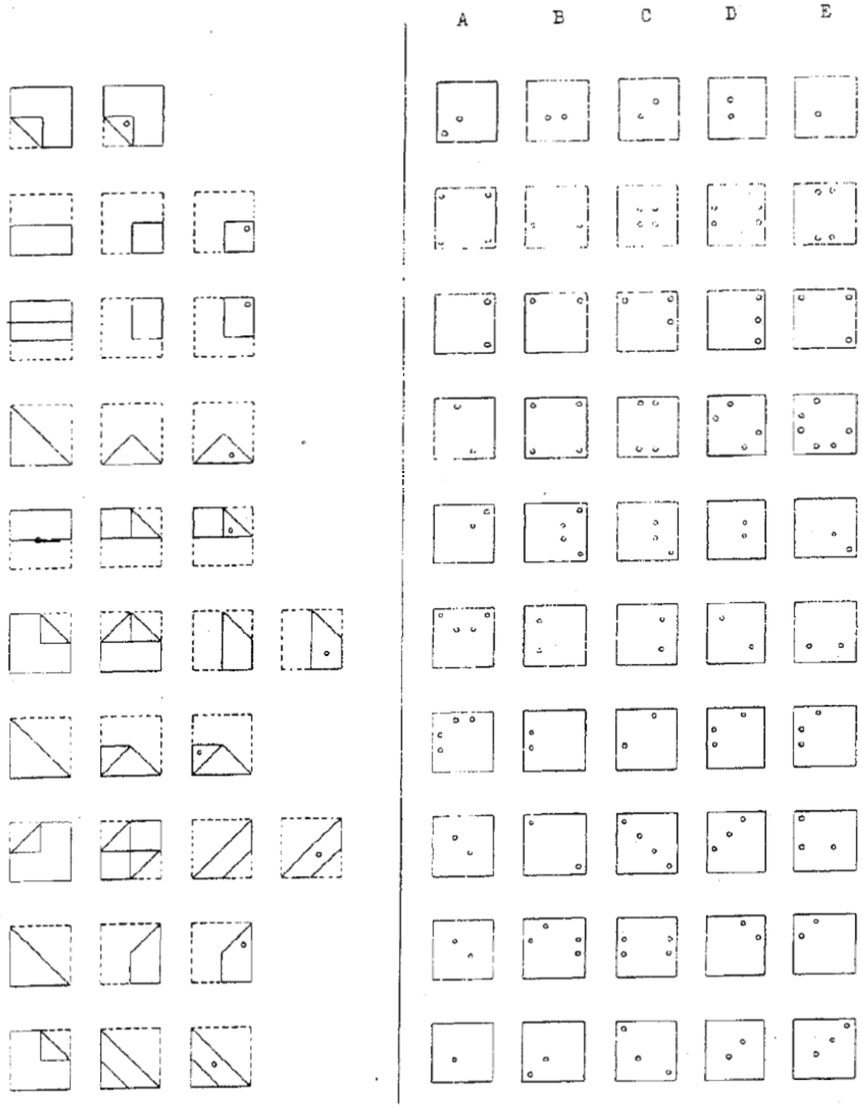
You will have 3 minutes for each of the two parts of this test. Each part has 1 page. When you have finished Part 1, STOP. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

Part 1 (3 minutes)

A B C D E

1 2 3 4 5 6 7 8 9 10



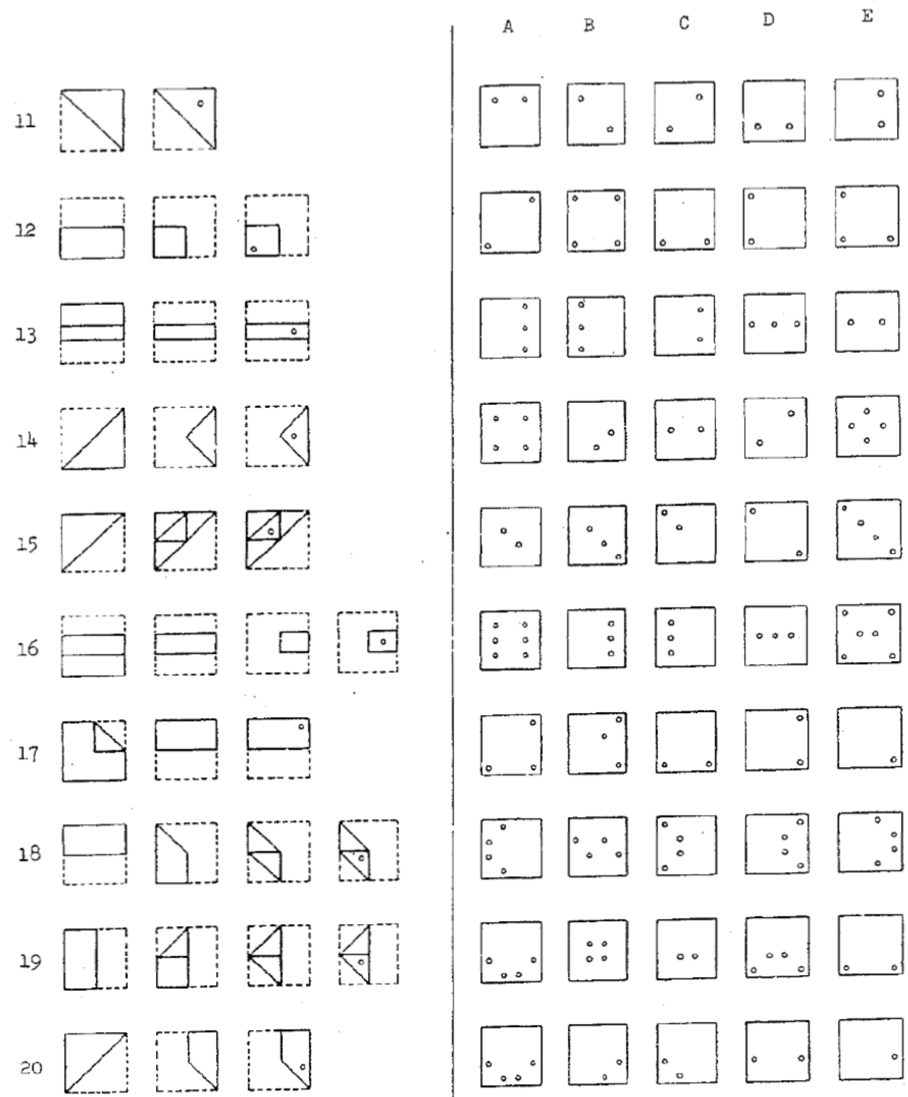
DO NOT GO ON TO THE NEXT PAGE UNTIL ASKED TO DO SO.

STOP.

ISU IRB # 1 09-386
Approved Date: 28 September 2009
Expiration Date: 12 September 2010

Page 3 VZ-2

Part 2 (3 minutes)



DO NOT GO BACK TO PART 1, AND

DO NOT GO ON TO ANY OTHER TEST UNTIL ASKED TO DO SO. STOP.

APPENDIX F. EKSTROM ET AL. (1976) TEST OF VISUAL MEMORY

(continued on next page)

MV MEMORY, VISUAL

Factor

The ability to remember the configuration, location, and orientation of figural material

There has been considerable debate as to whether or not this factor is due to test content. Thurstone (1946) thought that "the memorizing factor transcends the nature of the content" but more recent research has demonstrated the existence of iconic memory, which is used to store visual impressions. This suggests that visual memory is not simply the result of test content but involves cognitive processes different from those used in other memory factors.

There may be sub-factors of visual memory. Guilford describes six figural memory abilities. Petrov (1970) has found separate factors both for iconic memory and for short-term retention of visual material.

Identification: Guilford, MFU, MFC, and MFR, possibly others.

References: 21, 33, 55, 86, 91, 108, 109, 155, 164, 165, 174, and 179.

Building Memory -- MV-2

The subject is asked to indicate the location of a number of buildings seen on a previously studied map.

Length of each part: 12 items, 4 minutes for memorizing,
4 minutes for testing

Suitable for grades 6-16

ISU IRB # 1 09-386
Approved Date: 28 September 2009
Expiration Date: 12 September 2010

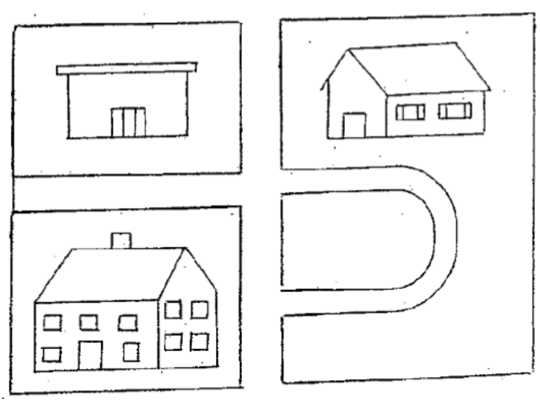
Name ____________________

BUILDING MEMORY -- MV-2

This is a test of your ability to remember the position of things on a street map.

You will be given a map with streets and buildings and other structures to study. After you have had some time to learn the street layout and the different kinds of structures, you will be asked to turn to a test page. On that page you will find the street map and numbered pictures of some of the structures. You will be asked to put an x on the letter that shows where each of the structures was located on the study map.

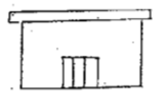
Now look at this simple and enlarged sample:



After you have studied the sample above for a minute, turn to the next page.

Look at the numbered houses on the left. For each item mark an X on the letter below each building that corresponds with where each house was located on the study map.

1.



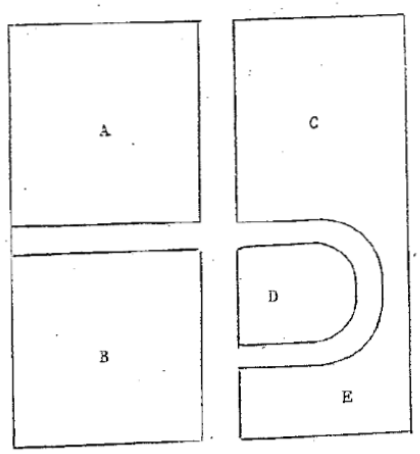
A B C D E

2.

A B C D E

3.

A B C D E



Your answers for sample item 1 should be A, for 2, C, and for 3, B.

Your score on this test will be the number of buildings placed correctly minus a fraction of the number wrong. Therefore, it will <u>not</u> be to your advantage to guess unless you can eliminate some of the locations as definitely wrong.

There are two sections to each part of this test. The first section is a map which you will study for <u>4 minutes</u>. The second is the test section and contains 12 structures to be located on the map. You will have <u>4 minutes</u> to mark your answers. Mark A, B, C, D, or E for each building. In the test section, the buildings will be mixed up and not necessarily near the part of the map where you first saw them.

This test has two parts. When you have finished Part 1, STOP. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO

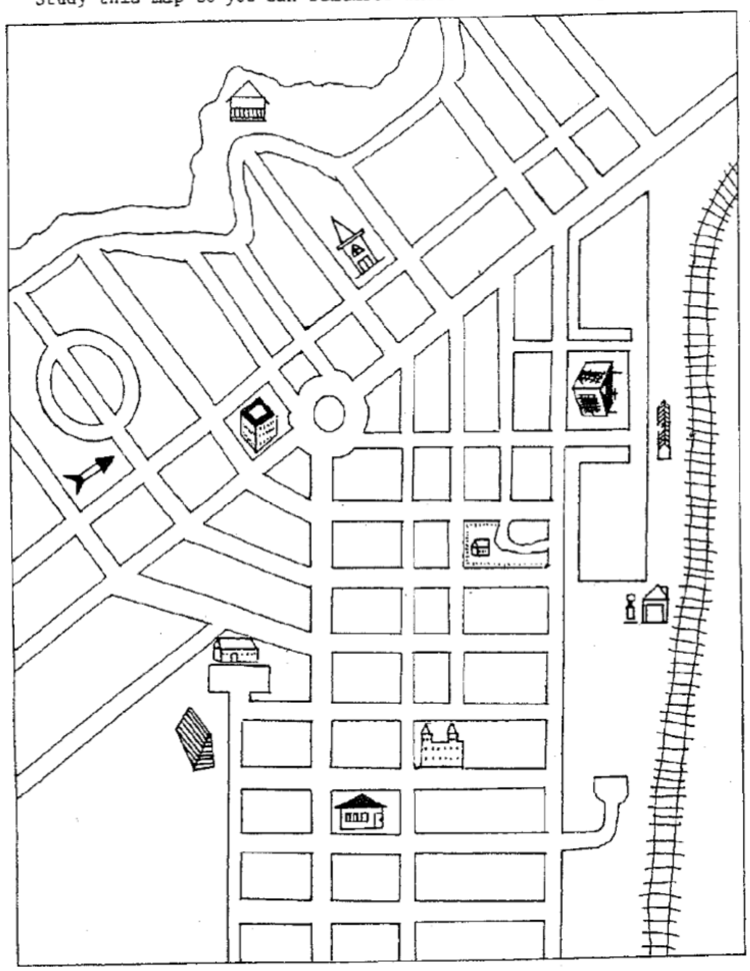
MV-2

Page 3

STUDY PAGE

Part 1 (4 minutes)

Study this map so you can remember where each building is located.



DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO

STOP

Page 4 MV-2

TEST PAGE

Part 1 (4 minutes)

Mark an X on the letter below each building that shows where it was seen on the map.

1. 
A B C D E

2. 
A B C D E

3. 
A B C D E

4. 
A B C D E

5. 
A B C D E

6. 
A B C D E

7. 
A B C D E

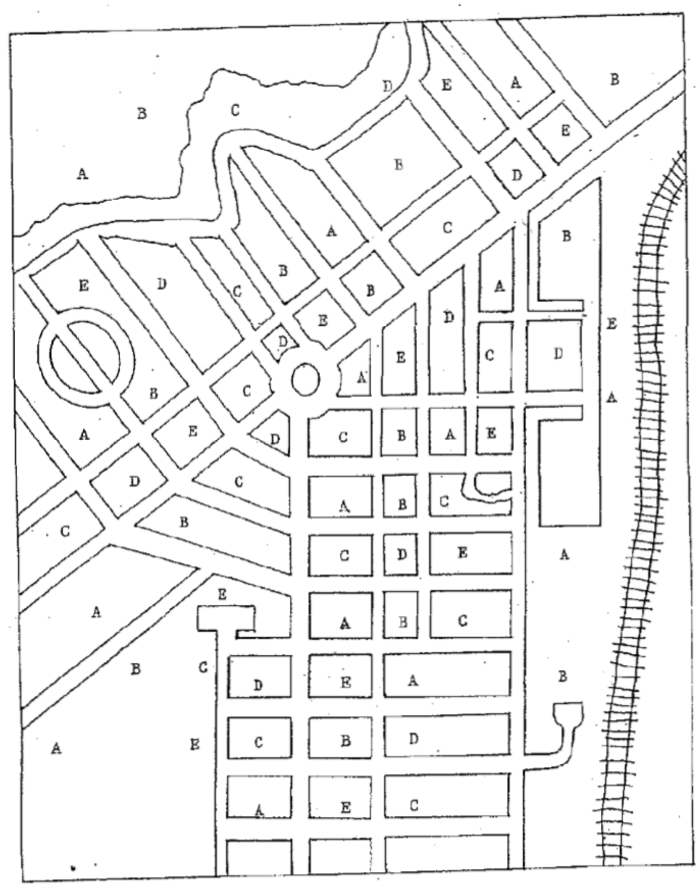
8. 
A B C D E

9. 
A B C D E

10. 
A B C D E

11. 
A B C D E

12. 
A B C D E



DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO

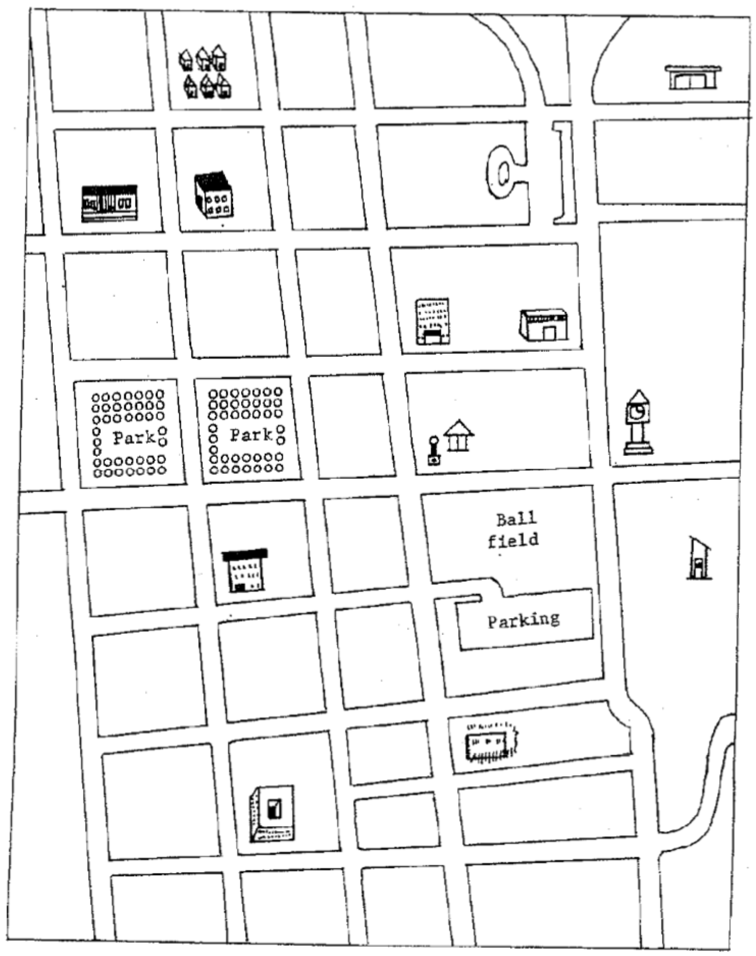
STOP.

ISU IRB # 1 09-386
Approved Date: 28 September 2009
Expiration Date: 12 September 2010

STUDY PAGE

Part 2 (4 minutes)

Study this map so you can remember where each building is located.



DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO

STOP.

TEST PAGE

Part 2 (4 minutes)

Mark an X on the letter below each building that shows where it was seen on the map.

13. 

A B C D E

14. 

A B C D E

15. 

A B C D E

16. 

A B C D E

17. 

A B C D E

18. 

A B C D E

19. 

A B C D E

20. 

A B C D E

21. 

A B C D E

22. 

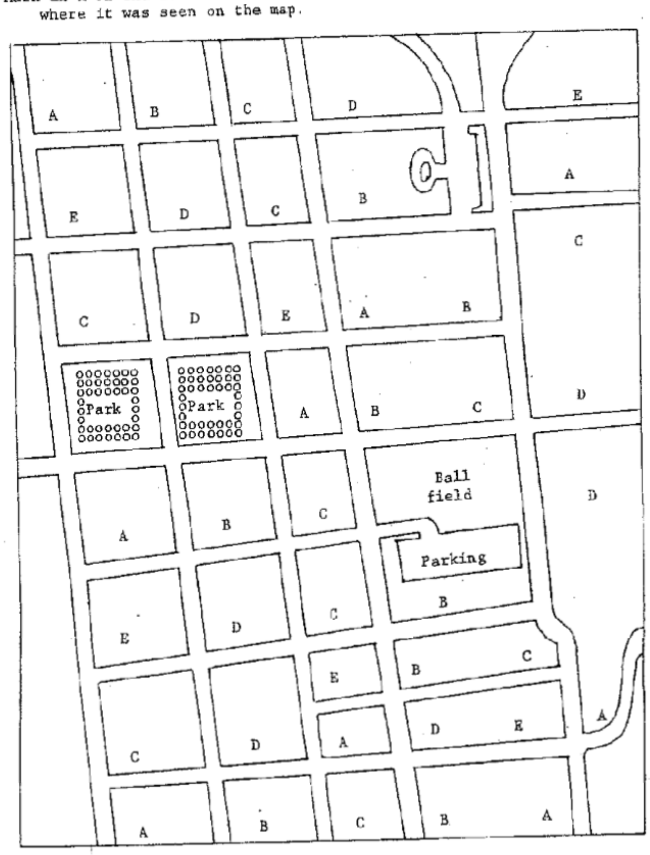
A B C D E

23. 

A B C D E

24. 

A B C D E



DO NOT GO BACK TO PART 1 AND DO NOT GO ON TO ANY OTHER TEST

STOP.

APPENDIX G. KOZHEVNIKOV ET AL. (2006) TEST OF PERSPECTIVE-TAKING ABILITY

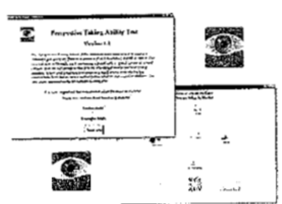
(continued on next page)

ISU IRB # 1 09-388
Approved Date: 28 September 2009
Expiration Date: 12 September 2010

Perspective Taking Ability Test (PTA-Test)

Recent research of Prof. Maria Kozhevnikov (Department of Psychology, George Mason University, VA) has shown that there are two distinct abilities: mental rotation (an ability to imagine rotation of objects from a fixed perspective) and perspective taking (an ability to imagine a reoriented-self) [pdf]. The second skill (perspective-taking) is the skill, which is important for navigating in space.

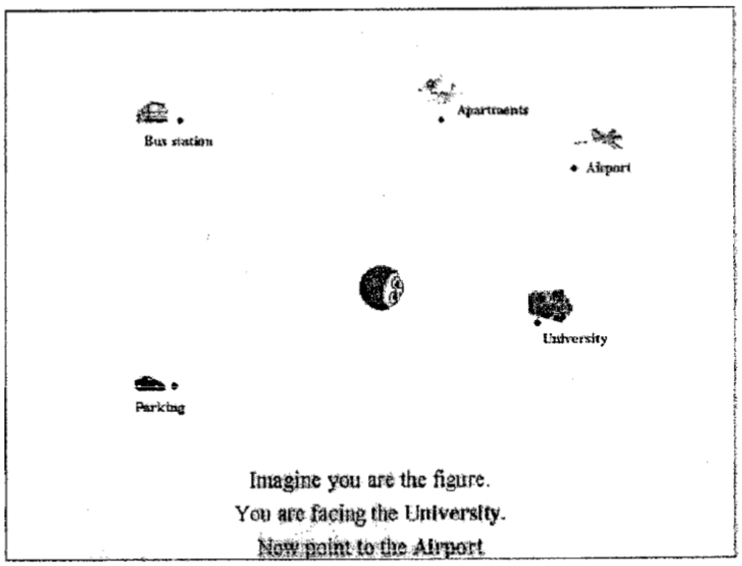
Up until now, the existing tests did not dissociate successfully between mental rotation and perspective-taking abilities, since most existing tests could be solved by using mental rotation as well as perspective taking strategy. As a result, all the existing commercially-available spatial tests measure mostly mental rotation ability (e.g., the ability to imagine rotating objects from a fixed perspective), which is a different ability not related to navigational skills.



We have developed the Computerized Perspective-Taking Ability (CPTA) test to measure spatial orientation ability. This new test was successfully validated and copyrighted jointly by MM Virtual Design, LLC and Rutgers University (see PTA test features). The results suggest that while solving this test, people in fact encode the objects shown on the display with respect to a body-centered coordinate system. It was also shown that while this test predicts reliably the spatial navigational abilities, mental rotation tests do not.

Our new Computerized Perspective-Taking Ability Test is the first valid measure of spatial orientation ability and could be successfully used for research as well as for training purposes and personnel selection in the professions that require high navigational abilities (e.g., astronauts, pilots, drivers).

ISU IRB # 1	09-386
Approved Date:	28 September 2009
Expiration Date:	12 September 2010



APPENDIX H. PAPER MAP STUDY: LIST OF CODES IN THINK-ALOUD PROTOCOLS

# pattern recognized	turn signaled
address expected	acquire cue
address verified	added street name
anchored on landmark	attempting to plan
check address	change target
check map	check address
check surroundings	express frustration
confused	identified map relation erroneously
target location estimated	intermediate goal set
edit map	intersection expected
street expected	landmark expected
heading selected	landmark identified
identify street	learn about area layout
landmark added	looking for street sign
map relation identified	near-target address verified second time
map rotated	numbering pattern recognized
navigation plan reinforced	orient self with regard to cardinal directions
near-target address verified	realize going the wrong way
noted address location	recalled travel sequence
planning	select intermediate target
planning check map	target address error found
position located on map	
recall solution to previous scenario	
recall target	
recognize # pattern	
select heading	
select target	
signal turn	
street identified	
target address error suspected	
target changed	
target recalled	
target selected	
target sequenced	
target street identified	
tried to locate target but failed	

APPENDIX I. PAPER MAP STUDY: POST-STUDY QUESTIONNAIRE

(continued on next page)

ISU IRB # 1 09-386
Approved Date: 28 September 2009
Expiration Date: 12 September 2010

FIELD EXERCISE TESTS & QUESTIONNAIRE

Direction Test

A test of direction will occur just before the field exercise begins. The answer will be recorded by facilitator.

1. "Point due North."

Field exercise has been administered and is now complete.

Starting Point Test & Direction Test

A test to determine if the participant can locate their starting point will be administered followed by an additional test of direction.

1. "Point to your starting location."
2. "Point due North."

Post-Questionnaire

Questionnaire questions will be read to the participant. The answers will be recorded and later transcribed

1. How did you decide what address to start with?
2. How did you decide what address to do next? Did you change this approach for later addresses?
3. What features of the map were most helpful?
4. What features do you wish were on the map?
5. What features of the map could you have gone without?
6. Did the setting affect your approach to completing the tasks (e.g. weather, traffic, etc.)? If so, how?
7. How hard was it for you to find the addresses on the ground? Very Easy [1] – [5] Very Difficult
8. What address was the easiest to verify?
9. What address was the most difficult to verify?
10. What distracted you from the task?

APPENDIX J. STATIONARY SIMULATION STUDY: ADVERTISING

(continued on next page)

Newspaper advertisement

Participants Needed for Tablet PC Study

We are looking for participants who are familiar with computers to perform tasks using tablet PC software

Those that participate in the study will be offered compensation

Contact Michelle Rusch at mlrusch@iastate.edu or (515) 294-9773

Email notification

Participants Needed for Tablet PC Study

We are looking for participants who are familiar with computers to perform tasks using tablet PC software.

Those that participate in the study will be offered compensation.

For more information, please contact Michelle Rusch at mlrusch@iastate.edu or (515) 294-9773.

RESEARCH STUDY

Participants needed for a study on software interface design for the tablet PC.

We are looking for participants who are familiar with computers to perform tasks using tablet PC software. If you volunteer to participate, your performance using a user interface will be captured.

The experiment takes about two hours. Subjects who participate in the study will be provided with compensation.

For more information or to schedule an appointment, please contact:

Email: mlrusch@iastate.edu

(515) 294-9773

Michelle Rusch

Research Study Michelle: 515-294-9773 mlrusch@iastate.edu	Research Study Michelle: 515-294-9773 mlrusch@iastate.edu	Research Study Michelle: 515-294-9773 mlrusch@iastate.edu	Research Study Michelle: 515-294-9773 mlrusch@iastate.edu	Research Study Michelle: 515-294-9773 mlrusch@iastate.edu	Research Study Michelle: 515-294-9773 mlrusch@iastate.edu	Research Study Michelle: 515-294-9773 mlrusch@iastate.edu	Research Study Michelle: 515-294-9773 mlrusch@iastate.edu	Research Study Michelle: 515-294-9773 mlrusch@iastate.edu	Research Study Michelle: 515-294-9773 mlrusch@iastate.edu

APPENDIX K. STATIONARY SIMULATION STUDY: INFORMED CONSENT FORM

(continued on next page)

Letter of introduction with elements of consent
Software interface design

Our goal is to study software interface design on tablet PCs.

The exercise will take about two hours to complete. During the study you may expect the following procedures to be followed. You will complete two short questionnaires, a few training and cognition assessments, and one exercise using a software interface.

There are no known risks to participation other than ergonomic concerns that are normally associated in the usage of the tablet PC.

There are no direct benefits to you as a participant other than the educational experience of being involved in an experiment. Your participation is helping the United States Census Bureau to develop better software.

By participating in this study, you will be provided with a $30 gift card for your participation.

Your participation in this study is completely voluntary and you may withdraw from the study at any time.

Records identifying participants will be kept confidential. Results will be released in summary form only.

Only researchers from Iowa State University working on this project will have access to data collected during this study.

The data collected in this research may be used for educational or scientific purposes and may be presented at scientific and/or educational meetings or published in professional journals. Published results will be in summary form only with no personal identifying information.

For further information about the study contact Michelle Rusch or Dr. Sarah Nusser at (515) 294-9773. If you have any questions about the rights of research subjects or research-related injury, please contact Janice Canny, Director, Office of Research Assurances (515) 294-4566, jcs1959@iastate.edu.

Participant's Name (printed) ______________________________

______________________________ ____________________
(Participant's Signature) (Date)

INVESTIGATOR STATEMENT

I certify that the participant has been given adequate time to read and learn about the study and all of their questions have been answered. It is my opinion that the participant understands the purpose, risks, benefits and the procedures that will be followed in this study and has voluntarily agreed to participate.

______________________________ ____________________
(Signature of Person Obtaining Informed Consent) (Date)

ORC 05/05

APPENDIX L. FIELD AND VR STUDY: ADVERTISING AND SCHEDULING SCRIPTS

(continued on next page)

PARTICIPANTS NEEDED

for ISU study.

Participants must be 18 or older, fluent in English, and capable of using a mobile device.

Compensation will be offered.

For more information or to schedule an appointment, please contact:

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

GEORGI BATINOV
Email: batinov@iastate.edu
Phone: (515) 450-5435

Newspaper advertisement

Participants Needed for Research Study

Participants must be 18 or older, fluent in English, and capable of using a mobile device.

Participants in the study will be offered compensation.

For more information or to schedule an appointment, please contact **Georgi Batinov:** batinov@iastate.edu or (515) 450-5435.

Email advertisement

Subject: Participants Needed for Research Study

Participants must be 18 or older, fluent in English, and capable of using a mobile device.

Participants in the study will be offered compensation.

For more information or to schedule an appointment, please contact Georgi Batinov**:** batinov@iastate.edu or (515) 450-5435.

SCHEDULING SCRIPTS

The Doodle Online scheduling tool will be used – http://www.doodle.com

After a student has been scheduled, a follow up email will be sent with appointment information

Screening – Email

Subject: Address Verification Study - Screening

Dear [*Participant's Name*],

You have received this email because you have expressed an interest to participate in our Address Verification Study. The next step is to schedule you for the screening portion of the study. The screening will involve a background information questionnaire and a series of cognitive tests. Screening will take approximately 1 hour to complete, after which you will receive a $10 gift certificate as compensation.

You may be selected after this screening to later participate in a field exercise that will take approximately 2 hours to complete. We ask that you schedule this initial screening if and only if you intend to participate in the field exercise. You will receive a $20 gift certificate for participating in the exercise.

Click here [*Doodle Scheduling Link*] to schedule a screening appointment.

Georgi Batinov
Graduate Assistant
Department of Computer Science

Your participation in this study is completely voluntary and you may withdraw at any time. The data that we collect from you will be kept confidential. If you have any questions or concerns about this study, please contact Georgi Batinov@ 515-450-5435 or Dr. Les Miller @ 515-294-7934.

Screening – Phone

- "Hello, my name is [Scheduler's Name]. I am calling you because you have expressed an interest to participate in our Address Verification Study. I'd like to remind you that your participation in this study is completely voluntary and you may withdraw at any time. The data that we collect from you will be kept confidential."
- "May I continue?"
- "The next step is to schedule you for the screening portion of the study. The screening will involve a background information questionnaire and a series of cognitive tests. Screening will take approximately 1 hour to complete, after which you will receive a $10 gift certificate as compensation. You may be selected after this screening to later participate in a field exercise that will take approximately 2 hours to complete. We ask that you schedule this initial screening if and only if you intend to participate in the field exercise. You will receive a $20 gift certificate for participating in the exercise.
- "May I email you the information that I have discussed along with a link that will allow you to schedule your screening appointment?"
- "Thank you for your time."

SCHEDULING SCRIPTS (continued)

Field Exercise – Email

Subject: Address Verification Study - Field Exercise

Dear [*Participant's Name*],

You have received this email because you participated in the screening portion of our Address Verification Study. We have reviewed your screening information and would like to invite you to participate in the field exercise portion of the study. This exercise will take approximately 2 hours to complete. You will receive a $20 gift certificate for participating.

Please refer to your *Informed Consent* document for additional information regarding the study. You are welcome to contact us if you would like to receive another copy.

Click here [*Doodle Scheduling Link*] to schedule the field exercise.
Note: We may contact you to reschedule if weather conditions are not favorable.

Georgi Batinov
Graduate Assistant
Department of Computer Science

Your participation in this study is completely voluntary and you may withdraw at any time. The data that we collect from you will be kept confidential. If you have any questions or concerns about this study, please contact Georgi Batinov @ 515-450-5435 or Dr. Les Miller @ 515-294-7934..

APPENDIX M. FIELD AND VR STUDY: INFORMED CONSENT FORM

(continued on next page)

INFORMED CONSENT

The purpose of this study is to find out how individuals survey addresses in the field.

You will complete a brief questionnaire followed by 4 cognitive assessments. This will take approximately 1 hour. Some subjects will be contacted at a later date for a field exercise. You will be given a $10 gift certificate for participation in this phase of the experiment, and you will need to sign a receipt for the gift certificate.

If selected for the field exercise, we will explain the task and train you. You will then be transported, via CyRide, to an Ames neighborhood where you will practice the procedures. Next, we will give you a list of addresses to find and verify against a map that we provide on a handheld computer. This exercise will take approximately 2 hours. There are no known risks to participation other than concerns that are normally associated with walking through a neighborhood. You will receive a $20 gift certificate for participating, and you will need to sign a receipt for it.

There are no direct benefits to you as a participant other than the educational experience of being involved in an study. Your participation is helping us learn how we can improve address listing methods.

Your participation in this study is completely voluntary and you may withdraw at any time. You may skip any part of this study that makes you feel uncomfortable or withdraw from the study at any time without penalty or loss of benefits to which you may otherwise be entitled.

Records identifying participants will be kept confidential to the extent permitted by applicable laws and regulations and will not be made publicly available. However, federal government agencies, the National Science Foundation, and the Institutional Review Board (a committee that reviews and approves human subject research studies) may inspect and/or copy your records for quality assurance and data analysis. These records may contain private information. To ensure confidentiality to the extent permitted by law, the following measures will be taken:

Data that identifies participants will be kept confidential. The information taken from this session will be assigned a unique code. Your name will not be associated with this information. Only researchers from Iowa State University working on this project will have access to data collected during this study. Study records will be kept confidential under password protected computer files. Data will be retained for two years and then will be destroyed.

The data results from this research may be used for educational or scientific purposes and may be presented at scientific and/or educational meetings or published in professional journals. Results will be released in summary form only with no personal identifying information.

For further information about the study contact Georgi Batinov at (515) 450-5435 or Dr. Les Miller at (515) 294-7588. If you have any questions about the rights of research subjects, please contact the IRB Administrator, (515) 294-4566, IRB@iastate.edu, 1138 Pearson Hall, Ames, IA 50011.

Your signature indicates that you voluntarily agree to participate in this study, that the study has been explained to you, that you have been given the time to read the document and that your questions have been satisfactorily answered. You will receive a copy of the written informed consent.

Participant’s Name (printed)____

____ _____________________________

(Participant’s Signature) (Date)

INVESTIGATOR STATEMENT

I certify that the participant has been given adequate time to read and learn about the study and all of their questions have been answered. It is my opinion that the participant understands the purpose, risks, benefits and the procedures that will be followed in this study and has voluntarily agreed to participate.

____ _____________________________

(Signature of Person Obtaining Informed Consent) (Date)

APPENDIX N. FIELD AND VR STUDY: TRAINING SCRIPT

(continued on next page)

Date: 2/4/2011 **Observer:** ____________ **ID:** ____________

TRAINING SCRIPT

I will give you a map and a list of three addresses.

Please determine if the addresses from the list are properly shown on the map. You may verify them in any order.

There are four possible outcomes for each address on the list:

1. The address on the list appears correctly on the map and in the neighborhood – **No Change**;
2. The address on the list does appear on the map but is located in a different location in the neighborhood - **Move.**
3. The address on the list does not appear on the map but does appear in the neighborhood - **Add**;
4. The address on the list does not appear on the map or in the neighborhood - **Delete**;

You may approach the task however you see fit.

The first three addresses are for training. While you work on them I can answer your questions. Please ensure you are comfortable with the task before we move to the main exercise.

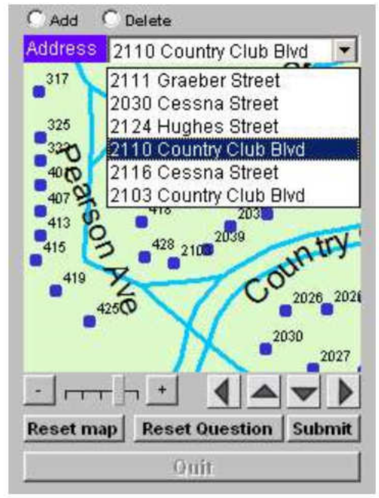
Practice Notes / Pointers

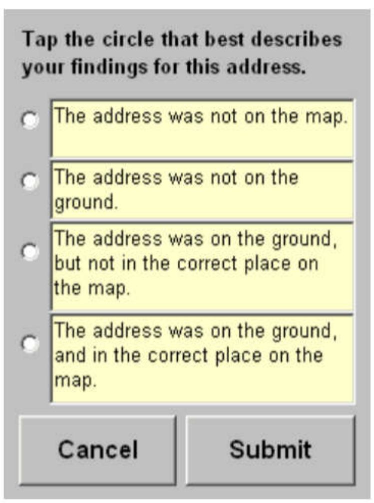
* did any strategies/observations emerge (e.g. numbering patterns) / note places where you can complement and coach participant *

APPENDIX O. FIELD AND VR STUDY: ADDRESS VERIFICATION SOFTWARE STORYBOARD\

(continued on next page)

MAP EDITOR STORYBOARD





APPENDIX P. FIELD AND VR STUDY OBSERVER CODING SHEET

(continued on next page)

CODING SHEET

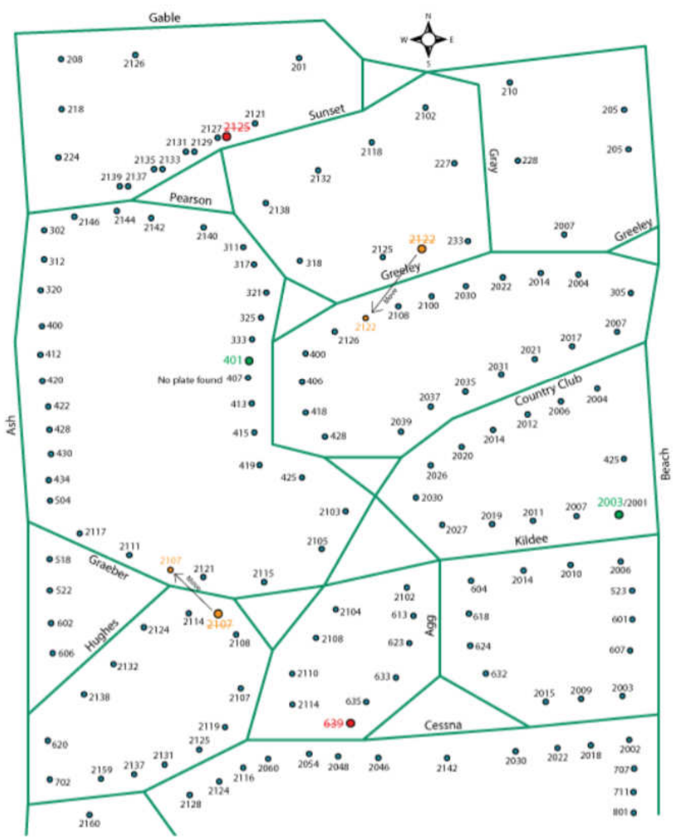
Date: 12/5/2015 **Observer**: ______________ **ID**: ____________

Address:______________ Order#:____ Action: **V A M D** Time of Verification:______

Map use frequency/style	**Body/Device Rotation**
Hesitation/Confusion	**Software Features Used**

ADDRESS NOTES

GENERAL NOTES



APPENDIX Q. STUDY COMPENSATION RECEIPT FORM

(continued on next page)

CONFIDENTIAL CONFIDENTIAL CONFIDENTIAL CONFIDENTIAL

Iowa State University
Research Participant Receipt Form (RPRF)
Use if this payment is less than $75

Iowa State University (ISU) is required to maintain the confidentiality of information about research study participants while still complying with record keeping requirements of the State of Iowa, the Internal Revenue Service (IRS), and funding agencies. The purpose of this form is to serve as documentation of the receipt of compensation associated with participation in a research study conducted by ISU personnel.

I, ______________________________, have received/or am requesting compensation in the form and amount indicated below:
(Print Research Participant Name)

☐ Cash $____________

☐ Check $____________

☑ Gift Certificate/Card $ 30.00

☐ Other Property – Describe: ________________________________

Value: $____________

______________________________ Research Participant Signature

______________ Date

TO ISU PERSONNEL:
Research participants may be given the opportunity to participate without receiving payment if they choose not to complete this receipt form.

This form provides documentation for gift certificates/cards or other property purchased by ISU p-card--keep original form as part of your p-card documentation.

If an ISU check needs to be issued for payment, attach RPRF to completed honoraria voucher and submit to Accounting, 3606 ASB.

APPENDIX R. PAPER MAP STUDY: INSTITUTIONAL REVIEW BOARD APPROVAL PAGE

(continued on next page)

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Research
Vice President for Research
1138 Pearson Hall
Ames, Iowa 50011-2207
515 294-4566
FAX 515 294-4267

DATE: 28 September 2009

TO: Kofi Whitney
226 Atanasoff Hall

CC: Dr. Les Miller
112 Atanasoff Hall

FROM: Roxanne Bappe, IRB Coordinator
Office for Responsible Research

TITLE: **How do we use a paper map? An exploratory study of spatial ability and decision making.**

IRB ID: **09-386**

Approval Date: 28 September 2009
Date for Continuing Review: 12 September 2010

The Chair of the Institutional Review Board of Iowa State University has reviewed and approved this project. Please refer to the IRB ID number shown above in all correspondence regarding this study.

Your study has been approved according to the dates shown above. To ensure compliance with federal regulations (45 CFR 46 & 21 CFR 56), please be sure to:

- **Use the documents with the IRB approval stamp** in your research.
- **Obtain IRB approval prior to implementing** any **changes** to the study by completing the "Continuing Review and/or Modification" form.
- **Immediately inform the IRB of** (1) **all serious and/or unexpected adverse experiences** involving risks to subjects or others; and (2) **any other unanticipated problems involving risks** to subjects or others.
- **Stop all research activity if IRB approval lapses,** unless continuation is necessary to prevent harm to research participants. Research activity can resume once IRB approval is reestablished.
- **Complete a new continuing review form** at least three to four weeks prior to the **date for continuing review** as noted above to provide sufficient time for the IRB to review and approve continuation of the study. We will send a courtesy reminder as this date approaches.

Research investigators are expected to comply with the principles of the Belmont Report, and state and federal regulations regarding the involvement of humans in research. These documents are located on the Office for Responsible Research website [www.compliance.iastate.edu] or available by calling (515) 294-4566.

Upon completion of the project, please submit a Project Closure Form to the Office for Responsible Research, 1138 Pearson Hall, to officially close the project.

APPENDIX S. FIELD AND VR STUDY: INSTITUTIONAL REVIEW BOARD APPROVAL PAGE

(continued on next page)

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Research
Vice President for Research
1138 Pearson Hall
Ames, Iowa 50011-2207
515 294-4566
FAX 515 294-4267

Date: 4/1/2010

To: Georgi Batinov
226 Atanasoff Hall

CC: Dr. Les Miller
112 Atanasoff Hall

From: Office for Responsible Research

Title: Do spatial ability differences persist in a virtual environment?

IRB Num: 10-075

Approval Date:	3/31/2010	**Continuing Review Date:**	3/30/2011
Submission Type:	New	**Review Type:**	Expedited

The project referenced above has received approval from the Institutional Review Board (IRB) at Iowa State University. Please refer to the IRB ID number shown above in all correspondence regarding this study.

Your study has been approved according to the dates shown above. To ensure compliance with federal regulations (45 CFR 46 & 21 CFR 56), please be sure to:

- **Use only the approved study materials** in your research, **including the recruitment materials and informed consent documents that have the IRB approval stamp.**
- **Obtain IRB approval prior to implementing <u>any</u> changes** to the study by submitting the "Continuing Review and/or Modification" form.
- **Immediately inform the IRB of** (1) **all serious and/or unexpected adverse experiences** involving risks to subjects or others; and (2) **any other unanticipated problems** involving risks to subjects or others.
- **Stop all research activity if IRB approval lapses,** unless continuation is necessary to prevent harm to research participants. Research activity can resume once IRB approval is reestablished.
- **Complete a new continuing review form** at least three to four weeks prior to the **date for continuing review** as noted above to provide sufficient time for the IRB to review and approve continuation of the study. We will send a courtesy reminder as this date approaches.

Research investigators are expected to comply with the principles of the Belmont Report, and state and federal regulations regarding the involvement of humans in research. These documents are located on the Office for Responsible Research website http://www.compliance.iastate.edu/irb/forms/ or available by calling (515) 294-4566.

Upon completion of the project, please submit a Project Closure Form to the Office for Responsible Research, 1138 Pearson Hall, to officially close the project.