The Pennsylvania State University
The Graduate School
College of Information Sciences and Technology

UNDERSTANDING HUMAN SPATIAL NAVIGATION BEHAVIORS: A NOVEL COGNITIVE MODELING APPROACH

A Dissertation in
Information Sciences and Technology

by

Changkun Zhao

© 2016 Changkun Zhao

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

May 2016
The dissertation of Changkun Zhao was reviewed and approved* by the following:

Frank E. Ritter  
Professor of Information Sciences and Technology  
Dissertation Advisor  
Chair of Committee

John Yen  
University Professor of Information Sciences and Technology

Xiaolong (Luke) Zhang  
Associate Professor of Information Sciences and Technology

Richard A. Carlson  
Professor of Psychology

Mary Beth Rosson  
Professor of Information Sciences and Technology  
Associate Dean for Graduate and Undergraduate Studies

*Signatures are on file in the Graduate School
ABSTRACT

Spatial navigation behavior is a basic ability for humans and animals to survive on the earth, as it allows us to seek food, return home, and localize friends. It is widely accepted that human navigation relies on some solid representations of space. Previous studies also show that there are two basic spatial representations: 1) the configurational (map) representation that consists of distances, Cartesian (absolute) directions, and geometric relations; 2) the sequential (route) presentation that involves landmarks and orientation sequences. However, how humans apply the two representations in their daily activities and the key factor of navigation process is still under debate. Two contradictory understandings are: 1) that configurational representation is a solid representation that could lead to accurate navigation behavior; 2) humans basically rely on the sequential representation, and the configurational representation can only lead to inaccurate navigation results.

This dissertation explored these issues with a new empirical scenario and a novel cognitive modeling approach. First, I conducted an empirical study using NavMaze to examine three new influence factors of navigation process: spatial retention, navigation preference, and mental rotation ability. The empirical results suggest that spatial retention is not a key factor or human navigation process, and the navigation performance is more correlated to navigation preference and mental rotation ability. This result reveals that human navigation process more replies on procedural skills. Second, I implemented a comprehensive cognitive model named NavModel in ACT-R to replicate empirical data. NavModel consists of a text-based testing platform for ACT-R, a mental rotation model based on an extended imaginal module of ACT-R, and an implementation of spatial representations and navigation strategies in ACT-R. The model fits the empirical data well; the mental rotation model, especially, can generate a very accurate prediction. In the modeling and data fitting process there are three new understandings of human
navigation process: 1) humans might rely on the sequential representation during navigation; 2) mental rotation ability is a key procedural skill in navigation; 3) humans use object separation and visual matching in the mental rotation process rather than rotating the entire object in their imagination.
# TABLE OF CONTENTS

LIST OF FIGURES ........................................................................................................ viii

LIST OF TABLES ........................................................................................................ x

ACKNOWLEDGEMENTS ................................................................................................. xi

Chapter 1 Introduction ................................................................................................. 1
  
  Dissertation overview .............................................................................................. 2
  Preview of contributions ......................................................................................... 4
  Preview of chapters ............................................................................................... 5

Chapter 2 Literature Review ....................................................................................... 7
  
  Introduction ............................................................................................................... 7
  Spatial cognition ..................................................................................................... 7
    Human spatial perception .................................................................................... 8
    Spatial memory ................................................................................................... 10
    Summary ............................................................................................................. 17
  Navigation ............................................................................................................... 18
    Navigation strategy ............................................................................................. 18
    Human navigation preference .......................................................................... 20
    Spatial studies in a virtual environment ........................................................... 21
    Summary ............................................................................................................. 22
  Cognitive architecture ............................................................................................ 22
    The Soar architecture ......................................................................................... 24
    ACT-R ................................................................................................................. 28
    Psychology studies with cognitive modeling .................................................... 35
    Navigation models based on cognitive architectures ......................................... 36
    Summary ............................................................................................................. 41
  Summary of literature review ................................................................................. 41

Chapter 3 A Preliminary Exploration on Modeling Navigation Behavior .................. 43
  
  Implementations of spatial memory representations .............................................. 43
    Sequential knowledge ......................................................................................... 44
    Configurational knowledge .............................................................................. 45
  Implementations of navigation strategies ................................................................. 46
    Preliminary results ............................................................................................. 47
    Implementing NavModel 1.0 .............................................................................. 49
    Summary ............................................................................................................. 51

Chapter 4 Empirical study: Discovering human navigation patterns and factors .......... 52
  
  Introduction ............................................................................................................ 52
  NavMaze ............................................................................................................... 53
Chapter 5 A Modified Cognitive Model of Human Navigation Process

Criteria of judging a cognitive model ................................................................. 79
Comparing navigation model with empirical data ........................................... 81
VIPER .............................................................................................................. 82
Model settings ................................................................................................. 83
Comparison of model prediction and empirical result .................................... 85
Discussion ...................................................................................................... 87
Mental rotation model .................................................................................... 88
Mental rotation model description ................................................................. 88
Comparing mental rotation model with empirical data .................................. 93
Adjusted mental rotation model .................................................................... 99
Comparing adjusted mental rotation model with the empirical data ............. 102
Discussion .................................................................................................... 106
Integrating the mental rotation model in navigation model ......................... 109
Comparing extended navigation model with empirical data ......................... 112
Model settings ............................................................................................... 112
Comparisons between model prediction and empirical result ..................... 113
Summary ....................................................................................................... 114

Chapter 6 Conclusion ...................................................................................... 116
Empirical navigation study .......................................................................... 116
Mental rotation model ................................................................................... 117
Navigation model .......................................................................................... 118
Contributions of this work ............................................................................ 119
An online platform to navigation experiment ............................................. 119
A study examining influential factors on navigation process ....................... 120
A cognitive model of mental rotation task ................................................... 120
A cognitive model to complete a complex navigation task ......................... 121
Limitations and future directions of this work .............................................. 122
Empirical study .............................................................................................. 122
Mental rotation model .................................................................................. 125
VIPER and NavModel communication ......................................................... 126

Appendix A Instruction letter for the navigation task .................................... 128
Appendix B Mental rotation test images ................................................................. 130
Appendix C List of memory retention test questions .............................................. 131
Appendix D Navigation preference scale ............................................................... 134
References .............................................................................................................. 135
LIST OF FIGURES

Figure 2-1. Baddeley’s model of working memory .......................................................... 11
Figure 2-2. Example of a question from Shepard and Metzler’s mental rotation experiment ................................................................. 16
Figure 2-3. An illustration of Soar’s basic components ................................................... 26
Figure 2-4. A brief history of ACT-R ............................................................................. 29
Figure 2-5. Schematic illustration of the current ACT-R architecture ......................... 31
Figure 2-6. An illustration of Gunzelmann and Lyon’s proposal on spatial competence ...... 33
Figure 2-7. An example diagram of conversion rules for Soar ........................................ 40
Figure 3-1. An example of a route chunk ..................................................................... 44
Figure 3-2. An example of a map chunk ...................................................................... 45
Figure 3-3. The influence of activation noise on the strategy running time ..................... 48
Figure 3-4. The influence of the activation noise on the running success rate of three navigation strategies ................................................................. 49
Figure 3-5. The high-level structure of NavModel v1.0 .................................................. 50
Figure 4-1. A screenshot of NavMaze ........................................................................... 54
Figure 4-2. The NavMaze layout that is used in the empirical study ............................... 56
Figure 4-3. An illustration of the experiment schedule .................................................. 58
Figure 4-4. An example of pre-survey to test participants’ spatial memory retention ....... 61
Figure 4-5. A screenshot of mental rotation test question on Qualtrics ......................... 63
Figure 4-6. Performance time of 4 different starting locations ....................................... 65
Figure 4-7. Comparison of two control groups on performance time ............................. 66
Figure 4-8. Effect of object rotation angles (clockwise) when number of blocks is 5 ....... 73
Figure 4-9. Response time of different number of blocks ............................................. 74
Figure 4-10. Response time of male and female participants ........................................... 75
Figure 5-1. NavModel experiment layout implemented in VIPER ........................................... 84
Figure 5-2. Experiment layout with 4 starting locations and the end location ...................... 85
Figure 5-3. Comparison between empirical data and NavModel 1.0 ................................. 86
Figure 5-4. The brief structure of mental rotation model v1.0 ................................................. 89
Figure 5-5. The logic flowchart of mental rotation model ......................................................... 91
Figure 5-6. Model fitting on rotation angles with default setting ........................................... 94
Figure 5-7. Model fitting on the complexity of rotation object with default setting ............. 95
Figure 5-8. Model fitting on rotation angles with best parameter setting ............................. 96
Figure 5-9. Model fitting on the complexity of rotation object with best parameter setting .. 97
Figure 5-10. The workflow of adjusted mental rotation model ............................................. 100
Figure 5-11. The brief structure of adjusted mental rotation model ................................. 102
Figure 5-12. Model fitting of adjusted mental rotation model on rotation angles .......... 104
Figure 5-13. Model fitting of adjusted mental rotation model on number of blocks .......... 105
Figure 5-14. An illustration of NavModel’s structure ......................................................... 110
Figure 5-15. Model fitting of total performance time of 4 starting locations ....................... 113
LIST OF TABLES

Table 4-1. Navigation experiment schedule ................................................................. 57

Table 4-2. Navigation strategy scale questions with exact questions, and their scale
analysis result ............................................................................................................. 69

Table 4-3. T-test results between the response time of 4 mental rotation angles ............. 73

Table 5-1. The criteria to judge a cognitive model ......................................................... 80

Table 5-2. A summary of response time per degree in existing mental rotation
experiments .................................................................................................................. 98

Table 5-3. Comparison of three versions of the mental rotation model ....................... 108
ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor, Dr. Frank Ritter. His kind support, guidance, and encouragement enabled me to finish this proposal. He is always willing to offer help to my research topics, as well as my personal life. Moreover, I appreciate the generous help and instructions from my committee members: Dr. John Yen, Dr. Xiaolong Zhang, and Dr. Richard Carlson. I would also like to thank my colleagues: Jonathan Morgan, Dr. Jaehyon Paik, Chris Dancy, and Ryan Kaulakis, who provided me with a great amount of useful comments. Lastly, I want to thank my beloved wife Na Li, my daughter Emma, and my parents.
Chapter 1

Introduction

Human spatial navigation is a basic and critical ability that is responsible for daily human activities, from driving to social interactions. It appears that humans have a very adequate ability to remember locations, to travel long distances, and to find their way out of mazes.

Consequently, scientists of psychology and education have been studying this topic for nearly a century (Cartwright & Collett, 1982; Foo, Duchon, Warren, & Tarr, 2007; Gallistel, 1990; Hart & Moore, 1973; Hirtle & Jonides, 1985; Kosslyn, 1980; Lynch, 1960; McNamara, Hardy, & Hirtle, 1989; Morris, 1981; O'Keefe & Nadel, 1978; Siegel & White, 1975; Stevens, 2014; Tolman & Honzik, 1930; Tolman, Ritchie, & Kalish, 1946; Wehner & Menzel, 1990). From their research, it is widely accepted that human navigation relies on some solid representations of space. Their studies also show that there are two basic spatial representations: 1) the configurational (map) representation that consists of distances, Cartesian (absolute) directions, and geometric relations; 2) the sequential (route) presentation that involves landmarks and orientation sequences.

However, how humans apply the two representations in their daily activities and the structure of these two representations has raised debates for decades (Bennet, 1996; Foo, Warren, & Tarr, 2005; Golledge, 1999). The main controversy of previous studies was the inconsistency between well-known spatial representations theories (Golledge, 1999) and the navigation pattern preferences that were revealed by recent experiments (Foo et al., 2007). Foo, Duchon, Warren, and Tarr’s experiment used a virtual environment to examine human navigation preferences, and their results suggest that humans basically rely on the sequential representation, and that the configurational representation typically leads to inaccurate navigation results. This result
contrasts the previous understanding of spatial representations, that the configurational representation is a solid representation that could lead to accurate navigation behavior.

Comparing previous human and animal studies of spatial cognition (Hart & Moore, 1973; Siegel & White, 1975; Tolman & Honzik, 1930), one main limitation of Foo and his colleagues’ work is that they used a simplified environment without multiple landmarks and geometric relations that aid humans to build up a configurational representation of the environment.

Based on the introduction above, it is clear that there are still some research gaps in understanding human navigation behaviors. Consequently, this dissertation tries to fill these gaps with a more comprehensive empirical study, and a new cognitive modeling approach.

**Dissertation overview**

To explain the controversy above and deeply understand human spatial navigation behavioral, I explored this question from two perspectives.

First, I conducted an empirical study to examine the effect of memory retention, navigation strategy preferences, and mental rotation ability on human participants’ navigation performance on a behavioral level. The empirical study was conducted in two main components that consisted of a controlled experiment spatial memory retention using the online navigation experiment platform NavMaze, and several surveys on navigation strategy preferences and mental rotation ability. The results showed that participants’ spatial memory retention is successfully manipulated but their navigation performance does not differ reliably. In addition, I found that their performance is highly correlated with navigation preference scores and mental rotation scores.

However, the behavioral analysis cannot reflect the internal cognitive mechanisms of navigation processes. In the second part, I propose a new approach to understand spatial
representations and navigation strategies by modeling human navigation behavior in a unified
cognitive architecture – ACT-R (Anderson et al., 2004). Cognitive Modeling is a promising
approach in human factors and psychology because it allows us to approximate complex
cognitive mechanisms by adjusting model variables to match sample human data (Sun, 2008).
The advantage of using ACT-R is that it provides a verified platform of human cognition,
memory, and especially visual attention. But the visual buffer is still not complete and needs to be
modified to some extent (Gunzelmann & Lyon, 2007).

Based on the findings in the behavioral analysis, the cognitive models focused on
modeling two processes: modeling mental rotation process, and implementing navigation
strategies and preference rules. To model mental rotation process, I introduced a mental rotation
model that integrates an extension of the imaginal module, visual model, and procedural learning
in ACT-R. The mental rotation model generates a good fit with the empirical data from my 2D
mental rotation experiment. During the modeling and data fitting process, I discovered a novel
explanation of mental rotation task: humans attend and rotate separate key features of the figure
rather than memorize and rotate the entire figure. With a well-fit mental rotation model, I further
developed NavModel, which involved a mental rotation model, implementations of spatial
memory in ACT-R’s declarative module, and implementations of navigation strategies and
preferences in ACT-R’s procedural module. The NavModel produces a decent fit with empirical
data, as it is the first ACT-R model that completes a complex navigation task. The good fit of
NavModel confirms Tar’s (2005) finding that humans apply fixed navigation preference on route-
based strategy. Additionally, the model also confirms that mental rotation ability plays an
important role in the navigation process.
Preview of contributions

In this dissertation, there are a couple of contributions for different research areas that range from psychology to computer science.

1) The mental rotation model provides a novel explanation of mental rotation process by combining mental imaginary and attention shifting together. This might reveal that attention shifting during mental rotation also plays a very important role.

2) This dissertation describes a cognitive model in ACT-R named NavModel, and it is one of a few ACT-R models that can complete a complex navigation task in a virtual environment. This model is reusable in multiple areas, including cognitive modeling, artificial intelligence, extreme events, and professional training.

3) This dissertation describes a mental rotation model in ACT-R by adding a mental rotation function that is verified by empirical data; it could be reusable for later models that involve the same mental rotation process. It can also be included in ACT-R as an optional extension to improve the competence of ACT-R.

4) NavMaze provides a highly compatible and reusable platform for all similar navigation studies. NavMaze is the only working navigation experiment system online for psychology. It has drawn some attention from the online learning and education community, and researchers from TellLab.

5) This dissertation introduces an empirical study based on NavMaze and Qualtrics. This study is highly reusable and more extendable than most traditional navigation studies.

6) The experiment result also shows the factors that influence navigation performance that deepen our understanding of human navigation process. The result suggests that spatial memory retention has no reliable influence, but mental rotation ability and navigation preference score is positively correlated with navigation performance.
7) This dissertation introduces a complete mental rotation model in ACT-R. This model is the first ACT-R model that completes a standard mental rotation task, and the prediction of this model is convincing.

8) This dissertation describes a cognitive model in ACT-R named NavModel, and it is one of a few ACT-R models that can complete complex navigation tasks in a virtual environment.

9) NavModel provides a novel explanation on human navigation strategy preference and adoption. Additionally, the model also shows how mental rotation ability plays an important role in the navigation process.

10) VIPER provides a highly editable environment. Theoretically, it can be modified for any layout and scenario because all layouts, objects, and actions are text-based and can be easily understood and edited. It is highly reusable for cognitive modeling, artificial intelligence, and network science studies.

**Preview of chapters**

I briefly preview the content of each chapter in this section. Chapter 2 is a literature review chapter that covers previous studies on navigation and spatial memory, and existing cognitive architectures. Chapter 3 introduces a preliminary cognitive model in ACT-R based on the understanding of previous literatures. The model is named NavModel. It also introduces a generic agent-based experiment platform called VIPER. Chapter 4 introduces a human subject study on the factors influencing navigation process, including an online experiment platform called NavMaze, the scenarios of experiment, and the result of the experiment. Chapter 5 introduces the process of improving NavModel to fit the empirical data. The NavModel 2.0 is then introduced, including a mental rotation model and new navigation preference mechanisms.
Chapter 6 is a conclusion chapter that includes general conclusion of this dissertation, lists of contributions, limitations of this dissertation, and future directions.
Chapter 2

Literature Review

Introduction

In this chapter, I summarize previous literature that relate to this dissertation. There are four areas of literature that will be covered in this review: 1) human spatial cognition that covers humans’ general spatial awareness and representations; 2) spatial memory that covers spatial working memory and long-term spatial memory; 3) spatial navigation that covers spatial navigation strategies and navigation preferences; and 4) cognitive architecture that covers popular cognitive architectures and existing cognitive models on navigation tasks in these cognitive architectures.

Spatial cognition

Human spatial cognition is a cognitive process that allows human to perceive, store, retrieve, and manipulate spatial knowledge. Human spatial cognition is an essential ability for humans because it is necessary for humans to navigate their surroundings, and is essential for space-related tasks like returning home or locating food (Wehner & Menzel, 1990). In the study of spatial cognition, researchers are particularly interested in two research topics related to spatial cognition: spatial knowledge and navigation strategies that utilize spatial representations. Spatial knowledge refers to a basic understanding of spatial geometry relations, spatial information cues to be identified, and event sequences in the spaces. Spatial navigation strategies refer to the decision patterns based on different types of spatial knowledge (Foo et al., 2005). It is widely
accepted that spatial knowledge plays a key role in human navigation, and the representations of spatial knowledge could also influence in navigation, the development of strategies, and navigation strategies shifting. This section will particularly focus on the existing theories on representations of spatial knowledge and navigation strategies from the perspectives of psychology and geography.

**Human spatial perception**

To build up the spatial representation of the environment, humans rely on their perception to perceive the world. Most types of human perception might contribute to building spatial representations, including visual perception, auditory perception, and haptic (touching) perception. Nevertheless, the most important perception source that humans use to build spatial representation is visual perception.

Visual perception has been long studied. There are a number of different views on visual perception in the field of psychology. Examples of these views include the Gestalt approach, the constructivist approach, and the ecological approach. (Sternberg, 2003)

In the constructivist approach, the visual perceiving process is suggested as a constructive process based on previous visual information. To perceive the world, humans need to reconstruct a model of the world by seeking, enhancing, and transforming existing visual materials. In the ecological approach, it is believed that the visual perception process is dynamic and it is highly interrelated with action and attention. Findlay and Gilchrist (2005) suggest that the eye chooses what to perceive in scenes based on what has been seen before and the intention of the seeing action. The Gestalt approach (Ritter, Baxter, & Churchill, 2014; Smith, 1988) suggests that a key function of visual perception is to separate the object from the background. Based on the theories of Gestalt psychologists, humans tend to select the simplest strategies to organize their perceptual
process. They summarize these strategies to several principles of perceptual organization, which are recognized as the Gestalt laws. They are five principles that organize perception information:

1) the principle of proximity, which suggests that close elements tend to be seen as an object; 2) the principle of similarity, which suggests that elements tend to be grouped together; 3) the principle of continuation, which suggests that humans tend to perceive continuous curves; 4) the principle of closure, which suggests that humans tend to fill in missing parts of an object; 5) the principle of common fate, which suggests that elements that move together will be perceived as the same object. (Ritter et al., 2014; Sternberg, 2003)

The psychology of 3-D space for humans, and thus also for cognitive models, can be divided into four spaces based on behavioral and neurological evidence (Previc, 1998). These are 1) peripersonal, near space that can be reached by hand; 2) focal extrapersonal, visual search and object recognition; 3) action extrapersonal, space they can directly and immediately navigate in, like a room, perhaps the space you can see; and 4) ambient extrapersonal, more distant space, like a football field to a city or country, where planning may be required to navigate. In each of these scales, spatial reasoning can also be performed, such as what is above what, and which is to the left of which (Corsi, 1972; Ritter et al., 2014).

The representations of these spaces can be done explicitly. 1) A checkerboard played by a robot hand that knows where all the pieces are with its eyes closed is an example of a local map, because pieces are represented. Knowing how to and simulating in your mind moving a toilet plunger from one shelf to a second is another example. 2) Near maps are manipulated with simple and direct walking, like being in a room. 3) Far maps are for larger spaces where time to travel and where the interpretation process (from map to action) will be important, and the space serves as a poor model of itself because of how big the space is and the need for abstraction.

The view of space could be reviewed with at least two other perspectives (Johnson, Wang, & Zhang, 2003). These perspectives include ego-based (egocentric) and plan-view based
(allocentric); and frame of reference, ego-based, body-centered, and object-based. These frameworks would provide a slightly different summary, but not substantially so, and the set used here provides a nice framework, particularly because after a brief review of displays, it allows more attention to be paid to larger scales.

Within so many levels and perspectives of space, humans exhibit a very good ability to perceive, understand, memorize, and reconstruct spatial information. This allows us to successfully achieve a lot of complex behavior in space like going home, maze navigation, or making shortcuts. Like verbal or procedure memories, researchers believe that spatial information is organized and memorized as some specific representations.

**Spatial memory**

Spatial memory is a part of the memory that is responsible for recording spatial information such as orientations and environment features. Based on their functionalities, spatial memory is usually categorized into two concepts: spatial working and long-term spatial memory. Spatial working memory is described as a limited capacity where human or animals can temporarily store and manipulate spatial information. In studies of vertebrates’ (mostly human) vision systems (Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999), spatial working memory is often considered as a subset of the vision system because it provides a temporary storage of the visual information before the brain processes it. Long-term spatial memory is the ability to recall high-level spatial information including layouts, routes, and environmental cues. Current research has found that long-term spatial memory is associated with a specific part of brain, the hippocampus (Moser & Moser, 1998; O'Keefe & Nadel, 1978). As the hippocampus is also known to be responsible for generating new memories for general experiences, researchers also believed that long-term spatial memory is highly associated with general long-term memory,
or declarative memory (Squire, 1992). It is widely accepted that long-term spatial memory exists in two formats, which are 1) cognitive map format with general layouts and landmarks, and 2) route format with episodes of egocentric spatial experiences (McNamara, 1986; Tolman, 1948). Nevertheless, how the brain would process long-term spatial memory and guide navigation when both formats of spatial memory are available is still raising a lot of discussions.

**Spatial working memory**

Spatial working memory is usually considered as a central component of the human working memory system. The first comprehensive model of working memory that was proposed by Baddeley and Hitch (Baddeley & Della, 1996; Baddeley & Hitch, 1974; Barrett, Eubank, & Marathe, 2006) counted spatial working memory (they called it “visual-spatial sketchpad”) as three major components of working memory. The original illustration of Baddeley’s working memory model is shown below (Baddeley & Wilson, 1985).

![Figure 2-1. Baddeley’s (1985) model of working memory.](image-url)
In Baddeley’s initial model (Baddeley & Hitch, 1974), the visual-spatial sketchpad has also been subdivided into two components. The first component is a visual cache that temporally holds and manipulates spatial images including shapes, colors, and verbal contexts. The second component is an inner sketchpad that handles spatial movement, and pass information into the central executive module. The phonological loop deals with an automatic process that collects audio information and temporarily stores the auditory sequences in the inner ears. The central executive is a central processor that binds all types of information resources, and schedules attention on selected tasks. In contrast to Baddeley’s working memory model, some researchers (Jones, Farrand, Stuart, & Morris, 1995; Taatgen & Rijn, 2010) view spatial working memory as a unified construct, where vision, spatial information, and verbal description are handled as levels of representation rather than storage of information. However, there are only a small number of researchers that hold this research perspective, and it still needs more supportive evidence.

**Long-term spatial memory**

Long-term spatial memory is the ability to recall high-level spatial information, including layouts, routes, and environmental cues. Current research has found that long-term spatial memory is associated with a specific part of the brain, the hippocampus (Moser & Moser, 1998; O'Keefe & Nadel, 1978). As the hippocampus is also known to be responsible for generating new memories for general experiences, researchers also believed that long-term spatial memory is highly associated with general long-term memory and declarative memory (Squire, 1992).

To better understand the structure of spatial memory, Tolman and his colleagues (Tolman, 1948; Tolman & Honzik, 1930; Tolman et al., 1946) studied the food finding and navigation behavior of rats, and they argued that a cognitive map in form of long-term memory exists in the brains of rats, as well as humans, to store spatial representation. Tolman (1948) also
illustrated evidence to support the existence of a cognitive map, which was an observation of novel shortcutting behavior of rats while navigating a maze. Although the original data from Tolman has been questioned, the notion of cognitive maps has gained wide acceptance. Since that work, numerous theories have been proposed to discuss the structures and representations of cognitive maps from the perspectives of spatial cognition, visual perception, and information processing. O’Keefe and Nadel (1978) explain the cognitive map as a neural structure in the hippocampus in the format of position vectors, and they also argued that a cognitive map should be the spatial relations acquired in an “unrewarded situation.” In contrast to a single representation of spatial knowledge, some researchers (Montello, 2001; Siegel & White, 1975) suggest a three-level representation with landmarks, route-based memory, and map-based memory.

The use of landmarks is the most fundamental element of spatial information, and it refers to an object or structure with prominent, distinct features in a large-scale environment. In navigation behavior, it is always used as a point of reference for localization, orientation, and route identification. Lovelace et al. (1999) categorized landmarks into four types: 1) the decision point, where the reorientation and self-localization are needed for navigation; 2) the potential decision point, where a reorientation is possibly needed for future navigation, but not for the current route; 3) the route marks that help to stay on the right path; 4) the distant landmarks that help to confirm distance between two points. Lynch (1960) suggests that landmarks are often used as an external reference rather than a node on the route. It could be any distinct points near the route, and the quality of a landmark depends on how it contrasts to the background. Despite those definitions, Raubal and Winter (2002) focus on the factors or attributes that cause an object to be selected as a landmark. They suggest that three types of attractions would determine the choosing of an object to be used as a landmark: 1) visual attraction, which indicates a visual contrast with background in term of color, facade area, shape, and visibility; 2) semantic
attraction that focuses on the distinct meaning of a feature, including cultural and historical importance, and explicit marks; 3) structural attraction, which refers to the feature with geometrical importance, such as interactions and boundaries.

The route-based representation is a sequential memory which consists of landmarks, egocentric orientation, distance and locations (Avraamides & Carlson, 2003; Bennet, 1996). Some researchers (Bennet, 1996; O'Keefe & Nadel, 1978) argue that route-based representation is a weaker form of spatial memory, because it is easily disturbed by changing landmarks or other spatial clues. In contrast, a series of recent studies (Foo et al., 2007; Foo et al., 2005) report that humans primarily follow familiar routes in navigation, which indicates the route-based representation is not a secondary representation, but rather a strong and sufficient form of spatial memory. In human spatial cognition, the route-based learning often takes place first when coming into an unfamiliar environment. As the route-based memory gathers, humans can gradually construct a map-based memory of the environment by recognizing landmarks, connecting positions, and changing egocentric coordinates (Hart & Moore, 1973).

The map-based representation involves a Euclidean map, including landmarks, exocentric orientations, and relative directions. Compared to a route-based representation, the map-based representation is considered the strongest format of spatial representation for two reasons: first, it principally could represent all spatial knowledge structure by encoding geometric relations between points accurately; second, it enables humans to achieve high-level spatial reasoning processes such as detouring and constructing novel shortcuts (Gallistel, 1990). Some researchers (Poucet, 1993; Tversky, 2003) extend the theory of map-based representation by discussing its hierarchical structure and scalability. They suggest that the map-based representation should have a scale system with hierarchical location structure in it to manage large-scale spatial information.

To combine the two theories, Jacob and Schenk (2003) suggested a parallel map theory to combine two types of representations. In their theory, they proposed a “local map” that represents
local spatial information as a middle state of two representations. The local map is created when a human is memorizing routes, and a global map representation will be built by combining several local maps.

**Mental rotation ability**

Mental rotation is the ability for humans to rotate mental representations to handle visual images. This ability is illustrated as a subsection of spatial memory because mental rotation ability is also considered as a sub-function of spatial working memory, and mental rotation is associated with the process of orientation in navigation process. This perspective has also been supported by the most recent neuroscience research that is boosted by using nuclear magnetic resonances (MRI) technologies.

Shepard and Metzler (1971) conducted the first research focused on examining the human visual imagination of manipulating image orientations. The original research method showed a set of representations of 3D objects to participants, and let participants decide whether this set of representations are rotations of the same object or not. An example question from their mental rotation experiment is shown in Figure 2-2.

The result showed that participants’ reaction time of making decisions was linearly proportional to the angle of the representations from the original position. The angle from the original position ranges from 0 to 180 degrees, meaning there is no difference between clockwise rotation and counter-clockwise rotation. Shepard and Metzler explained their elegant result that participants had a clear mental image and this image can be rotated at a steady rate. Based on their experiment settings, they concluded the rotation rate is 0.09s per degree, but this rate may vary under different experiment complexities. After the celebrated experiment, Shepard and his students (R. Shepard & Cooper, 1982) extended their research scope on mental rotation of
complex objects, such as objects with rigid edges, and their results remained very supportive of their original theories.

![Image of complex objects]

Figure 2-2. Example of a question from Shepard and Metzler’s (1971) mental rotation experiment.

Soon after Shepard and Metzler’s original research was published, some arguments have arisen on the existence of the mental imaginary that is responsible for manipulating and rotating object images. An alternative explanation from Just and Carpenter (Carpenter & Just, 1978; Just & Carpenter, 1976) on the linear relation between reaction time and rotation angle took account of the cost of eye movements when participants were tracking and comparing two pictures. The results also suggested that the total length of eye movements is nicely linearly proportional to reaction time. Their explanation (Carpenter & Just, 1978) was soon questioned by a study on visual
impaired participants. Those participants who were generally believed not to experience visual mental imagery still exhibited an ability to rotate haptic or kinesthetic images.

Controversy continues about the underlying mechanisms of mental rotation, and neuroscientists have tried to find out more with the assistance of fMRI. They gave participants the same mental rotation experiment, and they have found activation in Brodmann's areas 7A and 7B, the middle frontal gyrus, extra-striate cortex, the hand somatosensory cortex, and frontal cortex (Alivisatos, 1992; Cohen & Bookheimer, 1994; Tootell et al., 1995). Nevertheless, at the current stage, neuroscientists are still not able to support any theories of mental rotation mechanisms.

Summary

In this section, I mainly covered human spatial cognition from two aspects: human spatial awareness and spatial memory. In spatial awareness, I focused on introducing spatial perception and levels of human perception. With this in mind, I will be able to classify all navigation problems into different levels of spatial perception. In this dissertation, I will only focus on navigation problems on the level of a far map in explicit representations.

In spatial memory, I spent most of my effort on introducing two widely accepted understandings of long-term spatial memory. The route-based representation is a sequential memory that consists of landmarks, egocentric orientation, distance, and locations. The map-based allocentric representation is a high-level layout of spaces that requires a more complex transformation process. Based on my review, there is one question that still garners a lot of controversy: the exact format of the two representations and how they transfer across tasks.

I also introduced the mental rotation ability that is considered an important component of human spatial ability, and which has been nicely supported by empirical data. However, there are
still some arguments on the internal mechanisms of mental rotation, and the existence of mental imagery that stores and manipulates visual images.

**Navigation**

The study of navigation has a long history. The earliest work on this topic took place over half a century ago, when Tolman (1948) and Tolman and Ritchie (Tolman et al., 1946) conducted some of the first studies on this topic, and they coined the concept of ‘cognitive map’, that is, a mental map to present the spatial layout of the environment. Based on their study on rats, Tolman also hypothesized that cognitive maps correspond to sets of associations in the long-term memory of both humans and other mammals. O’Keefe and Nadel (1978) grounded Tolman’s assertion in a function analysis of a rat’s brain from a physiological perspective, arguing that cognitive maps are, in essence, sets of position vectors stored in pyramid cells of the rat’s hippocampus. Since then, some further studies on rodents (Lipp & Wolfer, 1998; McNaughton et al., 1996) and birds (Claydon & Dickinson, 1998; Sherry & Duff, 1996) have showed the same function in their hippocampi to store allocentric locations of environment. Another series of recent work on primates (Amaral, Insausti, & Cowan, 1983; Goldman-Rakic, Selemon, & Schwartz, 1984) discovered that animals with more complicated neural systems, such as monkeys, maintain their neural representation of allocentric information in the hippocampus as well as the posterior parietal cortex.

**Navigation strategy**

Based on all these findings of animal navigation research, studies on human navigation have been extended from neural functions to the cognitive and behavioral levels, such as stimulus
and manipulations that influence navigation, or the effect of gender or age. In addition, different from navigation research on animals, computer-aided virtual reality has been widely applied in order to study human navigation. Although it is not possible to replicate all spatial cues that humans could perceive in real world paradigms, such as lack of vestibular input, or narrower visual angles, it still has been shown that behaviors patterns (such as building up cognitive maps) applied in the virtual environment are comparable to those performed in the real world. However, it is hard to quantitatively compare participants’ performance between real world settings and virtual reality settings.

After decades of research in the real world and virtual reality, studies have categorized two different aspects of human navigation patterns that are supported by behavioral and neurological evidence. The first pattern is the allocentric strategy that corresponds to the representation of a cognitive map. As humans move through the environment, they maintain and constantly update their allocentric location, which consists of measures of distance and absolute directions (north, south, east, and west), based on internal and external spatial cues. When humans navigate to an unseen location, they locate the goal location and their current location in their allocentric map, then compute the egocentric distance and directions of the goal (Hart & Moore, 1973; Hirtle & Jonides, 1985; McNamara et al., 1989; Montello, 2001; Zhao, Morgan, & Ritter, 2013).

Contrary to the allocentric strategy, some evidence shows that humans also navigate relying on their memory on navigating processes that consist of egocentric movements and landmarks (Bennet, 1996; O'Keefe & Nadel, 1978). As a human moving along a route, he (or she) maintains an egocentric representation by updating an egocentric position vector that contains positions with landmarks, and associative movements (McNamara, 1986; McNamara et al., 1989; Sheltion & McNamara, 1997; Tarr, Bulthoff, Zabinski, & Blanz, 1997). Such an updating process is conducted by adding the displacement vector of the current egocentric view to the human’s
previous egocentric vector. Meanwhile, an allocentric representation of cognitive maps may be built up explicitly during the process, but it is typically incomplete and dynamic (Wang & Simons, 1999; Wang & Spelke, 2000; Zhao et al., 2013). Compared to allocentric maps, the egocentric representation is easier for humans to adopt because it can be used directly to guide humans’ self-locomotion by retrieving previous movements and computing a route from a set of egocentric vectors.

**Human navigation preference**

Based on these two widely accepted categories of navigation strategy, some further studies examined human preference on navigation strategies under various scenarios. I can roughly summarize these studies in two categories based on their experimental paradigms: 1) exploring and navigating in an unfamiliar environment, which is also referred to as a “way-finding” task; 2) navigating in a perfectly learned environment, which is also referred to as “memory-based navigation.” A typical setting for a way-finding experiment is to take participants into an unfamiliar environment, such as a complex or a town center, and ask participants to explore and navigate to a given position under manipulated conditions (Golledge, 1999; Ottosson, 1987). Plenty of studies on this issue have been done, focusing on influential factors (Cornell, Heth, & Alberts, 1994; Cousins, Siegel, & Maxwell, 1983; Lawton, 1994), and implications to improve way-finding performance (Kippel & Winter, 2005; Li & Kippel, 2012; X. Zhang, 2007).

Meanwhile, memory-based navigation has also gained plenty of interest, but the settings of those studies on memory-based navigation has varied extensively. The Morris Water Task (MWT) (Morris, 1981; Morris, Garrud, Rawlins, & O'Keefe, 1982) is a classic example that includes both declarative memory and procedural skills. MWT used mice as study participants and made them swim in a pool of water with only one location where they were able to stand.
This task was developed to examine spatial learning and memory on rodents, then, some later work (Astur, Ortiz, & Sutherland, 1998; Astur, Tropp, Sava, Constable, & Markus, 2004; Hamilton, Driscoll, & Sutherland, 2002) extended this task by implementing a computer-based version of the MWT to examine the relation between spatial learning and navigation behaviors on human participants. The task provides some strong evidence to support the notion that humans could construct and hold a cognitive map by using external location references to navigate. Another series of studies with completely different settings (Foo et al., 2007; Foo et al., 2005) also argue that map-based memory could not work without landmarks because landmarks provide a reference for humans to be aware of the relative position to the goal position. And they concluded that following a route is the only preference that their subjects applied. In sum, there are still many issues on memory-based navigation under discussion. Moreover, the effects of memory retention, which is a basic assumption for memory-based navigation, has not been systematically manipulated and measured.

**Spatial studies in a virtual environment**

In the recent years, researchers started to take advantage of virtual reality techniques to provide substitute experimentation in the real world, because virtual reality allows us to quantitatively manipulate the visual information during navigation behaviors. Tarr and Warren (2002) used a head-mounted display to simulate an interactive environment. Their work was an improvement over the pencil-and-paper tasks (Golledge, 1999) because it allowed us to record and accurately observe the navigation behavior.

Meanwhile, some other researchers (Hamilton et al., 2002) conduct experiments on desktop-based virtual reality. Hamilton et al. built a virtual Morris water task (VMWT) to replicate the Morris water task (MWT) (Morris, 1981; Morris et al., 1982), which is a classic task
to test spatial behavior on animals. Their VMWT is a desirable improvement of spatial cognition study because it allows human subjects to take MWT, and researchers are able to compare animal data with humans. Ruddle, Payne, and Jones (1998) extended desktop virtual reality to a large-scale environment, and their findings were that virtual reality could provide the same spatial information cues and orientation aids as the real world.

Summary

This section provided a general review of human spatial navigation studies in the past decades. It is clear that humans utilize two types of spatial information—route-based representation and map-based representation, in their long term memory to construct their knowledge of space. But few studies have been done to explain how the two types of spatial representations are organized and constructed in human long-term memory.

Based on the spatial representations, researchers suggest there are several navigation strategies to assist navigation behavior, and humans are able to switch navigation strategies in a complex situation, but the internal mechanisms of strategies selection are still not very clear, especially when both major long-term spatial memory representations are available.

Cognitive architecture

As a member of a science community, it should be common sense that unification is a good thing. A history of science could also be viewed as a history of the pursuit of theoretical unification: Euclid unified the theories of classical geometry; Newton unified the theories of classical mechanics; and Einstein tried to find the grand unified theory of physics in his later life.
At the beginning of the 1990s, Newell (1994) suggested that it is the right time to propose a unified theory of cognition to include all existing understanding of human cognition.

Newell explained the definition of unified theories of cognition in his book (Newell, 1994):

“A unified theory will unify our existing understanding of cognition. It will not be a brand-new theory that replaces current work at every turn. Rather, it will put together and synthesize what we know. …Its parts must work together… that permit specific cognitive theories to be inserted in them in some fashion.”

As Newell argued, despite the aim of unification, there are a couple of benefits to having a unified theory of cognition: 1) it helps to bring all individual theories and all constraints together to test behavior; 2) it could increase the identifiability of cognition theories; 3) it offers a chance to change cognitive psychology from discriminative science to an approximative style; 4) faster test will increase the rate of knowledge accumulation in this area; 5) it opens a way to cognitive applications (Newell, 1994).

Newell explained the definition of unified theories of cognition in his book (Newell, 1994):

“A unified theory will unify our existing understanding of cognition. It will not be a brand-new theory that replaces current work at every turn. Rather, it will put together and synthesize what we know. ……Its parts must work together…that permit specific cognitive theories to be inserted in them in some fashion.”

Newell, Rosenbloom and Larid (1989) also suggested that multiple theories of cognition was the most appropriate work to do instead of proposing a unique unified theory, because the approaches to implement a unified theory of cognition was still unclear. Consequently, his concern at the beginning was that multiple unified theories should be there for a while, and each implementation itself should be unified. In this case, researchers in cognitive science and
psychology would be able to compare among them and understand how to integrate cognitive theories in multiple ways.

To achieve a unified theory of cognition, Newell pointed out that cognitive architecture is a promising candidate for implementation because it has a compatible framework that not only could implement internal human cognition processes such as learning, memory, and reasoning, but also provides a general platform to insert new theories. With the development of cognitive architecture, researchers have summarized a couple of general characteristics of cognitive architectures (Laird, 2012; Newell, 1994; Newell et al., 1989), which would help us to better understand the definition of cognitive architecture.

1) It does not implement various aspects of human cognitive behavior, but rather human cognition as a whole.

2) It has robust behaviors with noise and uncertainties such as errors or mistakes.

3) It has a learning mechanism embedded, including declarative memory and skill acquisition.

4) It has a module to interact with the environment.

5) It should be parameter-free, which means this performance of the system does not depends on parameter adjusting.

The Soar architecture

The Soar is a cognitive architecture that was developed by Laird, Newell, and Rosenbloom (1987; Newell et al., 1989). The idea of Soar is based on the General Problem Solver (GPS) developed by Newell and Simon (1963), which is designed to simulate human general intelligence. Soar is based on a production system (rules should follow if … then…) that represents the basic rationales of problem solving. When it solves a problem, the Soar mechanism
searches the whole problem space that was built by a set of productions to archive a goal state. If the productions are competing or not matched accurately, the internal mechanism applies Bayesian reasoning (applied after Soar 6.0) to calculate a probabilistic solution for the goal state.

In the very early version of Soar, the decision making process is basically production matching with no previous experience. In version 5, Laird and Newell added a learning technique called *chunking* to transform the taken actions into the new rules. When the same situation is encountered, Soar could apply the new rules rather than going through the decision-making process again. In Soar, the *chunking* mechanism is implemented in two ways: chunking in motor control and chunking in working memory. In the motor control, Soar could build new rules to represent sequential actions. The mechanism only takes the inputs and outputs of the sequential actions and replicates the media actions as a whole process. In the working memory, Soar could merge memory nodes to build a hierarchical structure of knowledge, which may help Soar to manage large scale information with limited working memory (Laird, 2010, 2011, 2012).

In addition to the decision-making process, Soar also allows modeling communication with the external environment. Different from most cognitive architectures like ACT-R or CLARION (Sun, 2002), which uses individual motor and sensory modules, Soar applies a universal module to send and receive information to the environment called SML. SML stands for “Soar Makeup Language” and it was first introduced in Soar version 7. This module is implemented in Java and provides a general communication protocol that allows users to define their own communication messages to replicate sensory and motor control functions.
Challenges for Soar

Perception is one the main components of human cognition, and it handles all input information of humans. It represents all human sensory input, including vision, auditory, and touching. As I introduced in the section above, Soar does not have any specific modules to replicate vision, auditory, or other perception components. Instead, the Soar contributors proposed SML as a general solution for replicating sensory and motor control components. SML is based on XML, and it has many advantages:
1) A general solution of communication protocols provides an easier access to all perception and output motors. Users do not have to learn each perception module separately and consider their functional differences when modeling.

2) Technically, using XML is an easy and natural solution to develop and maintain SML. It also allows users to learn it and develop on it faster because most of the users are familiar with XML.

3) Users could embed the Soar kernel in their environment without launching the Soar debugger. This allows users to test their model much easier and faster.

4) It is a sufficient solution to support any type of perception information. However, the SML implementation’s major drawback is that it ignores the cognitive process of perception and motor control because SML assumes all I/O communications take place immediately with no cognitive costs. This drawback limits Soar’s ability to model some behavior that works with perception and motor control. For example, it is hard to model a cognitive process of visual recognition in Soar because it does not have an individual visual module and a theory of vision that would be embedded in it.

   Based on the current version of Soar, there are two possible solutions to overcome this major drawback. The first solution is to develop individual perception models, especially a visual module in the Soar core to support modeling spatial reasoning behavior. This is a very natural solution because a cognitive architecture is supposed to be an implementation of a unified theory of cognition, which should include all modules of human cognition as a whole system. However, this solution requires a lot of work because developers have to extend the core mechanism of Soar to allow its working memory process to internalize information from the perception modules.

   The second solution is to create an external perception module to hook up with SML. This solution is relatively straightforward because it does not require any changes of the Soar
core mechanisms and I/O. Developers only have to create a general representation of the space and reuse SML to communicate the spatial information with the Soar core.

Lathrop (2008) first attempted to create a visual module based on the second solution. He proposed a general representation of visual information based on SML named the Soar Visual System (SVS). One of the major challenges for Lathrop’s work was how to represent complex spatial and visual information by symbolic metrics and symbolic representation that could be processed in Soar. In Lathrop’s work, he suggested three distinct representations to support spatial and visual information:

1) Symbolic representation that represent general information of objects and relative spatial information.
2) Quantitative spatial representation that is generally used to represent Euclidean coordinate of objects and locations.
3) Visual depictive representation is from a privileged point of view and is constructed by the exact visual information pieces such as a bitmap of a tree or a building.

Each representation has trade-offs that depends on specific tasks. Based on the three basic representations, he (Lathrop, 2008) integrates SVS into the Soar architecture. This integration provides Soar a relatively complete access to spatial information, but it still needs further improvement because the integration is not fully functional. It utilizes external buffers of Soar. In the other words, SVS is still an external supplement that is not truly integrated with the Soar kernel.

**ACT-R**

ACT (Anderson et al., 2004; Anderson, Matessa, & Douglass, 1995) system is the accumulation of long-term work by Anderson, starting from HAM (Anderson & Bower, 1973),
an application of Anderson's thesis topic about semantic memory learning. The ACT system is created by integrating Anderson’s declarative memory learning theory with a production system theory of procedural learning. The first version of the ACT system was ACT-E (Anderson, 1976), and it became ACT* in 1983 by adding a subsymbolic level of cognition theory. By 1993, ACT-R emerged as a result of integrating rational analysis into the ACT system. Recently (2011), ACT-R has come to a new version 6.3.0 (Bothell, 2010a, 2010b), that includes more relevant theories and applications of cognition, such as a perception module, reinforcement learning, and instance-based retrieval (learning) (Gonzales & Lebiere, 2005). Figure 2-4 shows a brief history of ACT-R theory and versions.

Figure 2-4. A brief history of ACT-R (Anderson et al., 2004).
ACT-R is a cognitive architecture as well as a unified theory of cognition. It tries to provide a fully functional system that produces all aspects of human behavior in the cognition level. Even if the theories of minds consist of several parts and modules, such as perception, memory, skill, and reasoning, the ACT-R system is supposed to integrate all of them together (Anderson et al., 2004). To achieve this, ACT-R utilizes two basic mechanisms to implement human cognition as a whole. The first mechanism is a production system, which is also known as a rule-based system. Newell argued that the rule-based system is a plausible approximation of human reason processing, and it precisely replicates human procedural knowledge. In the production system of ACT-R, the unit of procedural knowledge is called productions (Bothell, 2010a, 2010b), which consists of a condition (or “IF” statement) to match current goal or sensory input and an action (or “then”) to execute the specific actions such as changing goal, modifying the state of declarative memory, and initiate motor actions. In the production system, I could merge any models by copying their code together. The models will run in parallel if there is no logical conflicts and competition between productions. The second internal mechanism is the buffers that temporarily store information in human working memory. Each of these buffers could hold a declarative memory chunk to be used or modified by the production system; for instance, the production system could send a request chunk in the manual buffer when it needs to perform the action. Among these buffers, a couple of them are particularly useful and important. They are the goal buffer, the retrieval buffer, two visual buffers, and a manual buffer. The goal buffer holds the goal state and keeps track of the reasoning process; the retrieval buffer holds the information that retrieved from declarative memory; the manual buffer holds control information for hand movement; the two visual buffers keep track of location of objects and identity of objects respectively. Although the buffer system covers a large proportion of human cognition structure, it is still far from being able to simulate all cognitive processes. However, the buffer system provides us an extendable platform, where I could insert any additional modules to facilitate our
models. The brief structure of buffers and modules in the ACT-R architecture is shown in Figure 2-5. In this figure, “red” represents internal modules, “blue” represents interactive modules, “white” represents buffers that associates with modules, “orange” represents core production system, and “green” represents all external environment that interacts with interactive modules.

![ACT-R architecture diagram](image)

Figure 2-5. Schematic illustration of the current ACT-R architecture. (Anderson et al., 2004).

ACT-R merges these two core mechanisms to implement human cognition as a whole framework, but not only because procedural memory and declarative memory are the most basic components of human cognition. There is another biological inspired theory under the structures of ACT-R. Anderson (Anderson et al., 2004) also concerns the correlation between ACT-R and human brain from the neurological science perspective, as he said:

"The basal ganglia and associated connection are thought to implement production rules in ACT-R. The cortical areas corresponding to these buffers projects to striatum, part of the basal
ganglia, which we hypothesize performs a pattern recognition function... the goal buffer associates with dorsolateral prefrontal (DLPFC)... the retrieval buffer is associated with the ventrolateral prefrontal vertex (VLPFC)... The other three modules/buffers (two visual buffers and a manual buffer) are all based on Perception Motor.”

Although the structure and supporting theories of ACT-R seem to be very plausible, it is still far from a validated cognitive architecture that is purposed to cover every aspect of human behavior. One of the main approaches to validate a cognitive architecture is by building cognitive models that match human behavior. A cognitive model is a replication of human process with the purpose of understanding and predicting human behavior. In contrast to cognitive architectures, cognitive models usually focus on a single high-level behavior or cognitive process, which could involve multiple modules of a cognitive architecture. A general cognitive modeling process involves three steps: comprehending tasks, modeling tasks in a cognitive architecture, and matching cognitive models to human data. A successful cognitive model could fit well with human data, and it could provide a precise prediction for human performance. Meanwhile, as cognitive models are built based on the theory of a cognitive architecture, a successful cognitive architecture provides a plausible validation to the cognitive architecture as well.

**Challenges for ACT-R**

As mentioned before, the ACT-R architecture is implemented by the contents of a set of buffers. Each buffer associates with a specific module, and serves as an information processing platform between modules and a production system. In the current version (6.0), ACT-R utilizes a visual module to process spatial information. However, the current visual module is built to represent computer-based objects, and it consists of a feature buffer to represent textures, attributes, and shapes of the objects; and a location buffer to represent objects’ coordinates on the
computer screen. Many researchers (Anderson, 2002; Anderson et al., 1995; Gunzelmann & Lyon, 2007) have argued that the current state of the ACT-R visual module does not provide the implementation of spatial competence because the realistic spatial representations are much more complicated than the computer-based visual space. Nevertheless, they believed that it is feasible to achieve spatial competence in ACT-R by adding a module and several buffers based on the current mechanisms of ACT-R. The approach they proposed consists of two parts: 1) adding an environmental buffer, an egocentric buffer, and spatial buffer to the current visual buffer to form a complete spatial representation in the three dimensional space (real world); and 2) implementing a spatial module to support several basic mechanisms of human spatial cognition including spatial transformation, spatial estimation, and magnitude calculations.

Figure 2-6. An illustration of Gunzelmann and Lyon’s proposal on spatial competence (2007).
Figure 2-6 shows the basic structure of Gunzelmann and Lyon’s (2007) proposal based on the current version (ACT-R 6.0) of the ACT-R structure. In the new spatial module they proposed, ACT-R could use a visual module to build up a solid representation of the environment as follows:

An extended visual module transformed the coordinate representation into an egocentric spatial representation that human cognition could process. The spatial module would combine and map the egocentric representation and deposit it into memory chunks in the declarative memory of ACT-R. These chunks could be either processed in the production system for decision-making or stored as long-term memory for later usage.

Although Gunzelmann and Lyon’s proposal provided a complete spatial representation and a spatial transformation mechanism to ACT-R, it still failed to explain some basic aspects of spatial perception, such as visual memory capacity that affects how many visual objects humans could perceive (Luck & Vogel, 1997); visual interference that represents how people tend to perceive objects with the same attributes (size, shape, or color) (Kosslyn, 1980, 1994).

Lyon, Gunzelmann and Gluck (2008) proposed a preliminary spatial model in ACT-R to implement visual capacity by simulating the mechanisms of declarative memory in ACT-R. Their model implementation consists of three parts of a theory, 1) visualized elements become less available with time; 2) activation is shared among similar elements; 3) visualized elements in nearby locations interfere with each other. The model could decently fit the human data in a path visualization experiment, and two parameters that are spatial interference parameter and activation threshold are proven as the influential factors on improving quantitative fit. However, this model is still a preliminary approach because they implemented their theoretical function by reusing the ACT-R declarative memory module rather than creating a separate spatial module to replicate spatial temporal/working memory. One of the major drawbacks is that the spatial
memory chunks and other declarative memory chunks have to share same memory parameters, such as decay speed and activation threshold.

**Psychology studies with cognitive modeling**

In a psychology study, one important approach to understanding internal mechanisms is modeling. This approach is widely accepted and used in the field of mathematical psychology and cognitive psychology because of three main advantages. First, modeling provides a promising approach to understanding human thought. As the human mind is a complex system, the understanding of the human mind is mainly from the observations of human behavior. Nevertheless, it would be extremely hard to understand the human mind in this way because the number of influential parameters is extremely large in a study. Such as Morris and colleagues (1982) and Tarr and Warren (2002) were only able to conduct a very restricted experiment to control variables.

Computational modeling is considered a new and more efficient approach because it is controllable and testable. The first advantage is that it allows researchers to simulate and test their hypothesis very quickly with the aid of a computer. The conventional psychology approaches, such as animal studies, observations, and questionnaires, usually take a long period of time to conduct, and they are sometimes unreliable if some confounding parameters exist. Computer-aided modeling could enable researchers to simulate their experiment quickly and precisely; in addition, it allows every parameter to be under control. Second, it allows psychologists to conduct some experiments that are impossible for conventional psychology approaches due to limited resources. For example, it was very hard to examine the psychological impact of the social network because such a study needs very large sample sets and human resources by conventional psychology approaches.
In general, psychology models could be categorized into computational models, mathematical models, and verbal-conceptual models. The computational models focus on implementing behavior process; the mathematical models replicate relations between parameters based on mathematical equations; the verbal-conceptual models describe structure, entities, and descriptions by natural language. In this section, I will particularly focus on summarizing existing computational models that represent spatial cognition and behavior, because my research interests focus on how to present human spatial behavior process in rational and memory level.

In regards to modeling tools, there are two types of cognitive computational models: 1) the general computational models that use general programming such as Java, C, or other mathematical tools; 2) another type of computational models that are based on cognitive architectures that implement a unified theory of cognition (Newell, 1994). Such cognitive architectures include Soar and ACT-R. The next section will focus on existing cognitive models for navigation processes.

### Navigation models based on cognitive architectures

Although many theories have been developed to understand human spatial memory and reasoning, we are still far from fully explaining all of these behaviors. One interesting approach for better understanding spatial reasoning and navigation is to develop cognitive models that seek to replicate the navigation behaviors observed in humans and animals (Fu, 2003; Kurup & Chandrasekaran, 2009; Lathrop, Wintermute, & Laird, 2010). There are a lot of models have been built to explain human spatial cognition based on cognitive architectures, especially ACT-R and Soar.

ACT-R is a cognitive architecture as well as a unified theory of cognition. It uses a production system to implement rule-based rationality, and numerous buffers to simulate human
perception. Specifically, ACT-R relies on a set of mechanisms (e.g., the utility mechanism and activation mechanisms) to simulate human memory, information processing, and reasoning. Gunzelmann and Lyon (2007) discuss the applicability of integrating a module for human spatial memory into ACT-R from both a theoretical and empirical perspective. They conclude that ACT-R can support modeling spatial cognition by providing mechanisms for learning and information processing. Gunzelmann and Anderson (2004) proposed a series of ACT-R models to execute a self-orientation task with multiple strategies. These models matched human data for this task. Based on these models, they then proposed a general extension of ACT-R 5.0 to complete a visual buffer and spatial module. In their proposed extension, they suggested adding two extra slots in the visual buffer to represent visual depth and visual textures. In addition, they argued the importance of an imaginary buffer in the human spatial behavior, and as a result they believed that mental imaginary would provide a tight connection between the visual buffer and spatial representation.

Researchers using ACT-R have not only modeled fairly localized tasks but also larger tasks such as navigating a maze, a community, or even a country. Fu (2003) developed a navigation model based on a 2D map representation implemented in ACT-R. By using a simple grid map as a testing environment, and coding the map representation into declarative memory, the model supported two basic navigation strategies: hill-climbing and a simple planning strategy. From Fu’s model and experiments, he suggested that humans would use a very direct and greedy navigation strategy such as hill-climbing rather than some complex planning strategies. More recently, Reitter and Lebiere (2010) presented a path-planning model that emulates the navigation strategies observed in humans using two components: a visual attention component and a spatial experience (memory) component. The visual attention component simulates the heuristics applied by human beings when performing a visual search in an unfamiliar environment. These heuristics include three basic tasks: straight-line extraction, visual
search, and goal recognition. The spatial experience component encodes into declarative memory the spatial information gained by the agent’s perceptual-motor module, generating a path or route representation. The agent then follows the path by activating declarative memory chunks sequentially. When there are existing paths available in declarative memory, the model will tend to navigate using the experience component. Otherwise, the agent tends to use the visual attention component.

Besides ACT-R models, there are a lot of models that have been used based on the Soar architecture. Different from ACT-R, Soar is a cognitive architecture that was built to simulate a human rational decision-making process and AI. It appears a similar rule-based production system to replicate human reasoning process, but Soar has a different approach to access human memory.

In the early versions of Soar, it only focused on the core mechanism of human information processing, but no perception module was considered. To interact with the environment, the Soar group developed a standard interaction protocol—“Soar Markup Language,” which is an XML-based interaction language. The SML is very helpful and efficient because it allows the environment client to send or receive command from the Soar client directly. However, this direct approach also neglects the key process of human cognition that is vision perception and spatial cognition. Lathrop and his colleagues (2008; Lathrop et al., 2010) proposed a general cognitive architecture named Soar/Visual System (SVS), which integrates symbolic (Soar), visual, and spatial representations. This architecture could also be recognized as a general extension of Soar with visual and spatial modules. The key component implementing this extension is that SVS implements a mental imagery to allow human cognition to move symbolic information among modules, including visual modules, spatial representation modules, and working memory. The visual module is designed to convert general visual information into symbolic information as the Soar cognition cores can only process symbolic information. In the
visual module, it utilizes a “visual predicate extraction” component to extract symbolic visual information, and pass it to the Soar working memory to construct a high-level spatial representation of the space. In this converting and extraction process, mental imagery module plays a very important role because it provides a temporal buffer to move information freely between visual module, working memory, and spatial representation.

As Soar is a popular cognitive architecture that has a long history, a lot of studies on spatial reasoning and navigation have been done based on it. Wintermute and Laird (2007, 2008) focused on integrating the symbolic representation of Soar with a diagram-like spatial representation. This bimodal spatial representation allows their motion model to better understand the environment and conduct a fine motion plan to navigate in the complex environment. Based on the bimodal representation of the space, Kurup and Chandrasekara (2009) proposed a model that could navigate in a large-scale environment. In their approach, they tried to build up a large-scale spatial representation to replicate the cognitive map that represents high-level spatial representation with landmarks, routes, nodes, and edges. To build up a bimodal representation, they proposed a general mapping diagram to convert general map information, including nodes, routes, and edges into symbolic spatial representation. Figure 2-7 shows an example diagram of their conversion rules. These conversion rules provide a general solution to build bimodal representation of the space, but one of the major drawbacks of their work is that it could only support map-based spatial information that is highly standardized and simplified.
Figure 2-7. An example diagram of conversion rules for Soar taken from Kurup and Chandrasekara (2009).

In this section, I introduced a useful approach to research complex human behavior that is modeling validated models via cognitive architectures. The advantages of this approach, such as low expense, high repeatability, and accessibilities, make it very popular in the cognitive science community. I briefly introduced two of the most popular cognitive architectures, the Soar and the ACT-R architectures. After evolutions over more than three decades, both architectures are capable to represent most aspects of human cognition, and hundreds of models were built via two architectures that deepen our understanding of human behavior. This section only selected a few promising models that are related to human spatial behaviors, but none of them explained how humans navigate and plan routes in real world scenarios.
Summary

In summary, cognitive architecture provides a promising approach towards a unified theory of cognition. Meanwhile, a cognitive architecture is also a computer application that enables researchers to test their cognitive theories rapidly. I introduced two mainstream cognitive architectures, Soar and ACT-R. Soar was built upon memory chunking theory and working memory stacking. It focuses on predicting behavior sequences and decision-making processes. ACT-R was built upon declarative activation theory and production rule matching. It focuses on predicting accurate reaction time and sequences of module activations that corresponding to different areas in human brain.

As this dissertation is interested in human navigation process, I summarized several cognitive models on navigation processes in both Soar and ACT-R. It turns out that there are only a few models on navigation, and most of them are still at a very early stage in development.

Summary of literature review

In this chapter, I covered three basic topics that are related to the interests of this dissertation. The first topic covered a basic understanding of human spatial cognition. Based on current research, it is clear that humans apply two types of long-term spatial memory, but the details of these two representations and their relation is still not clear.

The second topic was human navigation and navigation strategy. It is also clear that there are two types of strategies observed but only a little research and evidence has been reported on how humans apply these two strategies. In addition, how the factors affect navigation strategies are still not clear.
The third topic covered the history and popular examples of cognitive architectures. Although cognitive architectures have become a popular research method and direction for AI and psychology research, there is still a limited number of studies focusing on human navigation tasks that have been modeled.
Chapter 3

A Preliminary Exploration on Modeling Navigation Behavior

This chapter covers a preliminary implementation of spatial representations and navigation strategies, and this model is named NavModel (Tolman & Honzik, 1930). I did this preliminary work for two purposes: 1) to examine the feasibility of modeling spatial representations and navigation strategies in the declarative memory of ACT-R; and 2) to explore the cognitive differences between navigation strategies.

This model is different from previous navigation models (Best & Gerhart, 2011; Fu, 2003; Gunzelmann & Anderson, 2004; Lathrop, 2008; Reitter & Lebiere, 2010; Zhao, Hiam, Morgan, & Ritter, 2011). It implements spatial knowledge with multi-level structures and proposes three navigation strategies with different cognitive costs (Zhao et al., 2013).

More specifically, some work (Best & Gerhart, 2011; Fu, 2003; Zhao, Hiam, et al., 2011) only focus on the performance of the model by implementing some very efficient algorithm but are less cognitively plausible; some works (Reitter & Lebiere, 2010) lack multi-level representations of spatial memory; and some works (Gunzelmann & Anderson, 2004; Lathrop, 2008) lack implementation of navigation strategies and their shifting.

Implementations of spatial memory representations

To implement spatial knowledge in declarative memory, I first defined the basic location chunk that represents an individual waypoint in the environment with identification information such as objects, landmarks, and topological connections. This model used this chunk to construct both route knowledge and map knowledge.
**Sequential knowledge**

I used *route* chunks to represent sequential transitions between the start location and the end location. Different from the conventional route, the *route* chunk of this model only consists of 4 *locations*, and I implemented a longer path as a linked list of several *route* chunks. This approach was developed to match the limitations of human attention. When navigating along a long path, humans can only focus on a subset of the route because of the limitations of their working memory. According to the relevant studies (Luck & Vogel, 1997; G. Zhang & Simon, 1985), the average number of visual items that a human can hold in visual short-term memory ranges from 3 to 5. Consequently, I took the mid-number, 4, as the size of a *route* chunk. Finally, implementing a route with a list of subset route chunks enabled the model to integrate two long routes and also to discover shortcuts in the routes.

Figure 3-1 explains how to implement a long route with the route chunks. In this figure, I used 3 route chunks to implement a route consisting of 10-locations. The first route chunk contains the first 4 locations and an associative slot that points to the next route chunk.

![Figure 3-1. An example of a route chunk.](image-url)
**Configurational knowledge**

I used a *map* chunk to represent the configuration of the whole environment in a hierarchical structure. In this model, I created two chunk types to implement topological knowledge: (a) the *map* chunk to represent topological relations between locations; and (b) the *zone* chunk to implement hierarchical relations between locations. I created the *zone* chunk because humans use clustering to organize large amounts of map knowledge, again, because of working memory limitations. Thus, humans cannot, without some external aid, process all the spatial information of a large environment, such as a skyscraper or town. The node chunks of a *zone* chunk could be either *location* chunks or *zone* chunks, and I used an ordered tree algorithm (Hirtle & Jonides, 1985) to build up a hierarchical structure of the environment. Figure 3-2 shows an example of map chunk within a hierarchical structure. In this example, the highest chunk is the *Zone1* chunk that contains 5 sub zone chunks with topological relations between them. In the secondary level, the *Zone6* consists of 4 locations chunks, and these locations chunks could also be contained by other zone chunks, such as *Zone2* and *Zone4*.

![Figure 3-2. An example of a map chunk.](image)
Implementations of navigation strategies

Humans use multiple navigation strategies based on different types of spatial knowledge (Bennet, 1996; O'Keefe & Nadel, 1978). In this model, I implemented three navigation strategies:

(a) The route following strategy (or route-based strategy) is the most basic strategy. Based on sequential knowledge, humans will conduct sequential actions to follow landmarks in order. This strategy is considered weaker in the sense that referents depend on sequence, which if broken leaves no other cues. Humans, even when primarily using this strategy, generally supplement this learned order with other knowledge.

(b) The goal-directed strategy (or map-based strategy) is usually applied when the goal is visible or they are quite familiar with nearby environments. I implement the goal-directed strategy based on the configuration knowledge of the environment because map knowledge could provide sufficient information for orientation and finding a direct path to the goal. As I implement map knowledge in a hierarchical structure, the goal-directed strategy also plans a path hierarchically from the highest level to the lowest level. For example, if a person tries to navigate from NY to LA with the goal-directed strategy, I expect the traveler would plan starting from the state level and ending at the street level.

(c) The hybrid strategy implies some proceduralized skills that allow agents to identify regularities that then allow them to make predictions about the environment. Further, the prediction should be built based on a strong understanding of the environment configuration. More specifically, this strategy allows humans to achieve two high-level spatial behaviors: 1) route integration and 2) taking shortcuts. For route integration, it enables finding a path between
the end point of one route and the beginning point of another route; for shortcuts, it enables finding shortcuts.

As I noted earlier, humans show a general preference to select the cognitively least expensive strategies (Foo et al., 2007; Foo et al., 2005). More specifically, I interpret these strategy preferences in the following order: route-based strategy, the hybrid strategy, and map-based strategy. This preference, however, could be altered by individual difference and previous experience with successful shortcuts. My navigation model implements this preference as a proceduralized skill, and it could be changed by ACT-R’s reinforcement learning with each navigation strategy.

Preliminary results

To test NavModel, I ran some preliminary experiments to compare NavModel with some single strategy models, which are a subset of NavModel. I simplified the experiments by hardcoding all spatial information and predefined it in all three models.

Figure 3-3 shows the effect of adding noise to the process. Consequently, it is clear that all the curves increase as noise increases, meaning that activation noise disturbs the navigation process. The times for the hybrid strategy and map-based strategy are higher than the route-based strategy, suggesting that they are more cognitively expensive.

Comparing map-based strategy and the hybrid strategy, I noticed that the map-based strategy curve is higher than the hybrid strategy initially, but the difference disappears as noise increases. This is because the hybrid strategy uses route knowledge initially, and it gradually relies on map knowledge when the route-based strategy is no longer reliable under high cognitive noise. From this, I can conclude that, in terms of the retrieval cognitive load (equivalent to the complexity of the cognitive process) each strategy imposes:
map-based strategy ≥ the hybrid strategy > route-based strategy

Figure 3-3. The influence of activation noise on the strategy running time (the route line ends at 0.8 because all runs above that point fail).

Figure 3-4 shows the influence of activation noise on the strategy’s success rate. Increasing the activation noise decreases the success rate of the route following strategy most rapidly because some waypionts on the route could not be retrieved, and the route following strategy fails. This result matches empirical data (Bennet, 1996; O'Keefe & Nadel, 1978), that is the route knowledge is a weaker representation and the navigation strategy based on the route knowledge is easily disturbed.
Figure 3-4 also shows that map-based strategy and hybrid strategy can be disturbed by activation noise, but their failure rate is less sensitive. Finally, I could draw a conclusion that, in terms of robustness of the strategy to noise in cognition:

(Equation 3-2)

The hybrid strategy > map-based strategy > route-based strategy

Figure 3-4. The influence of the activation noise on the running success rate of three navigation strategies.

Implementing NavModel 1.0

Based on my basic implementations of spatial memory and navigation strategies, I started to put them together for a more complex navigation task. NavModel is designed to conduct
comprehensive navigation tasks, and simulate human navigation processes. Figure 3-5 shows a brief illustration of the NavModel structure above.

NavModel is a comprehensive ACT-R model that integrates spatial representations and navigation strategies that were introduced above. This model includes a shifting mechanism to switch navigation strategies to minimize the cognitive expense during navigation process. In a real navigation task, there are a lot of factors that are related to the cognitive expense, including visual process, memory retrieval, decision making, and motor control. However, some factors are fixed and could be not minimized by shifting navigation strategies; for example, controlling speed while driving is critical sub-task in navigation process but the cognitive load it requires could only decrease by practice and proceduralizing behaviors. Consequently, this model only focuses on memory retrieval and decision making process because spatial knowledge representation and navigation strategies play an important role in them.
Summary

In this chapter, I introduced a preliminary version of NavModel in ACT-R, which implements spatial knowledge with multi-level structures and proposes three navigation strategies with different cognitive costs. Based on previous literatures introduced in Chapter 2, this model implements two types of spatial representation in the declarative module of ACT; and it also implements three navigation strategies using production systems and utility mechanism of ACT-R. In addition, I also ran a test experiment on the navigation strategies, and found that the three navigation strategies: route-based strategy, map-based strategy, and hybrid strategy differs in term of cognitive costs and robustness. Lastly, I briefly introduced a preliminary version of NavModel to complete complex navigation tasks. More details about NavMaze will be discussed in Chapter 5.
Chapter 4

Empirical study: Discovering human navigation patterns and factors

Introduction

In this chapter, I present an empirical study on a human navigation task. The main purpose of this empirical study is to quantitatively test the hypotheses revealing the influential factors of human navigation processes. The second purpose is to collect empirical data for testing and validating the preliminary version of NavModel, and exploring possible directions to improve NavModel. This study is a relatively complex study that consists of two phases that are the learning phase and testing phase. It also consists of three sub-studies that are maze navigation study, navigation survey, and mental rotation test. The experiment procedures and materials are covered in this chapter. Some materials can be also found in the Appendix. After introducing the experiment, I also present the results of this experiment, and analyses of the results.

Experiment objectives

There are three objectives in the experiment:

1) Collecting human data to test and validate my ACT-R model on two navigation strategies.

2) Testing the hypotheses that were derived from my navigation strategy models, and previous literature.
3) The experiment includes an exploratory experiment that helps me observe and test how humans choose or switch navigation strategies.

In the experiment, I test three hypotheses:

H1: Greater spatial memory retention results in better navigation performance.
H2: Map-based strategy results in better navigation performance.
H3: Greater mental rotation ability results in better navigation performance.

**NavMaze**

I developed NavMaze in JavaScript and HTML with a computer graphic package “three.js.” I also developed a PHP program named “MazeTracker” that is included in the HTML file that can record user data, including key logs, response time, and all parameter settings. The recorded data sets are organized and stored in an Apache database. It can also export organized results and some preliminary analyses in “.excel,” “.csv,” or “.txt” format.

NavMaze is based on HTML and JavaScript because they are the mainstream programming language for webpage development, and they are compatible with all Internet browsers. Consequently, NavMaze can be accessed, reused, and modified by any user on the Internet. When I started to prepare this navigation experiment, the first obstacle I faced was that nearly all virtual reality experiments are not reusable because psychology researchers usually implemented their experiment by modifying existing 3D games. With the updates of games and supporting software, there are a lot of technology issues with experiments. In addition, as most of these experiments are modified from old games, it makes the experiment system very large and slow because the system has a lot of redundant running functions that are not related to the experiment. Compared to these heavy experiment systems, the core NavMaze is less than 1MB,
and it is accessible directly by Internet browsers, such as Firefox, Chrome, and Safari. In NavMaze, users can use the “up” key to move forward, use the “down” key to move backward, use the “left” key to turn left, and use the “right” key to turn right. The moving speed is set as 1000 simulated distance per second while the key is held down, with key sampling speed of 25 per second (in Figure 4-2, the size of maze is set as 15000*17000); the turning speed is set as 45 degrees per second.

Based on my knowledge and research, NavMaze is the only working experimental navigation system online. Right now it has drawn some attention from the online learning and education community, and researchers from TellLab\(^1\), which is an online psychology experiment platform that was developed by the Department of Psychology at Harvard, and is integrating NavMaze into their system. The source code of NavMaze is also available at Github for developers and researchers to download\(^2\).

![Figure 4-1. A screenshot of NavMaze (the number “17” in the middle of screenshot is the time spent in the game).](image)

\(^1\) http://lab.tellab.org/
\(^2\) https://github.com/blayer/3DMaze
Method

Materials and apparatus

In the experiment, I used a Mac computer to run a JavaScript-based virtual maze, NavMaze, under Firefox version 29.0.1. The display resolution 1920*1080 and the display was refreshed at a rate of 60 Hz with a display latency of 20ms. (Shown in Fig 4-1.)

In the virtual maze, there is a START location where all participants begin their navigation, and an END location where all participants navigate to. Additionally, there are 12 other landmarks located in the maze. These landmarks are all represented by animal figures and names, such as elephant, lion, or tiger. Figure 4-2 shows the layout that I used in this empirical study. In this figure, “red” indicates landmarks, “black” indicates start locations and end locations.
Figure 4-2. The NavMaze layout that I used in the empirical study (1, 2, 3, 4, indicate four start location in the testing session).

**Participants**

The participants used in these studies were students and employees at the Penn State University. The total number of the participants was 37, with 21 female participants and 16 male participants. Their age ranged from 18 to 25. One male participant was rejected from the experiment because he refused to follow the experimenter’s instruction to finish the learning session. The details of the participants and experiment schedule is shown in Table 4-1 with 36 participants in total.
Table 4-1. Navigation experiment schedule.

<table>
<thead>
<tr>
<th>Retention interval</th>
<th>Participant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
</tr>
<tr>
<td>1 day</td>
<td>8</td>
</tr>
<tr>
<td>7 days</td>
<td>7</td>
</tr>
<tr>
<td>Total participants</td>
<td>15</td>
</tr>
</tbody>
</table>

**Experiment design**

All participants were randomly assigned into two groups, and each group had different experiment settings. For the learning session, all participants of two groups were asked to explore the entire maze, and to sketch the layout. To make sure they memorized the entire layout and possible routes, they were asked to re-draw the entire layout based on their memory.

After that, participants in group 1 were asked to come back one day later, and participants in group 2 were asked to come back in 7 days. When they came back, they were asked to finish a new navigation task. In this task, all participants were requested to navigate based on their layout knowledge and distance cues to the goal location. Their performance data was collected to examine navigation strategy selection, as well as the relation between memory retention and navigation strategy.

After finishing the new navigation tasks, all participants were asked to take a survey on navigation preferences, and a mental rotation ability test. The details of experiment schedule is shown in Figure 4-3.
Figure 4-3. An illustration of the experiment schedule.

**Experiment procedure**

**Tutorial**

Before conducting the main experiment, all participants took a short tutorial to learn how to operate the environment. The participants played with the game using the keyboard in front of them and were asked to finish 4 basic tutorial tasks, including moving, turning, and using assisting information. This tutorial took 5 minutes in total, and the experimenter confirmed with every participant that they fully understood how to operate the software.
Learning session

All participants took the first experiment session right after they finished the tutorial. In this session, all participants were asked to take a learning task to learn the layout and possible routes in the maze.

Exploration

All participants begin navigation from the START location in the maze, and they are asked to explore the maze themselves to draw the layout of the maze and point out two possible routes between the END location and START location. A pencil and some pieces of paper were provided to the participants to sketch their results. Also, a list of all animals was presented to all participants to help them identify each landmark. Participants were given 10 minutes to explore and sketch themselves. After that, the experimenter gave them verbal instructions based on their drawings to help them sketch the layout. Some sample verbal instructions were as follows:

“Please go back the dog and elephant …okay …if you turn left here, you can see a new landmark …Then, please draw the location of this landmark.” or

“Can you check the list of animals, and tell me which animal you are missing? …Right. So, if you check the opposite side from the lion, you should be able to find a new landmark. …okay, can you sketch out how you got to this animal right now?”

During the exploration, NavMaze the trajectories of participants’ movement in the NavMaze including their positions and facing directions were recorded. The total time of exploration was also timed in minutes.
Checking the learning effect

When participants could successfully sketch the layout with all landmarks, they were asked to point out two possible routes between the START location and END location, and then to finish two trials of navigation along the two possible routes. If any mistakes happened while navigating along the routes, participants were asked to redo the trial.

All participants were asked to redraw the layout of the maze and landmarks independently with a pencil and a white paper. The list of landmarks is provided to help them. The experimenter would check their redrawing based on the layout structure and relative locations between landmarks to decide whether a participant passed the checkout or not. If a participant failed to redraw the layout, the participant would be given 2 minutes to skip the layout and retake the redrawing test. Participants would not proceed to further steps until they passed the redrawing test.

Memory test

After the manipulation check, participants were asked to take a short post-survey about the route and layout knowledge that they learned. The post-survey has 6 questions in total, 3 questions about route knowledge and 3 questions about layout knowledge. The order of the 6 questions was randomly assigned.

Session 2

Session 2 of the experiment was to examine the preference of navigation strategies when navigating in a familiar environment. To control participants’ memory retention, all participants
were randomly assigned with a retention interval between session 1 and session 2. The retention interval is designed to be 1 day, or 1 week. A more detailed schedule is shown in Table 4-1.

**Pre-survey**

The pre-survey of session 2 had the same format as the post-survey of session 1. It has 6 questions in total, with 3 questions about route knowledge, and 3 questions about layout knowledge. The order of these 6 questions was randomly assigned. An example question is shown in Figure 4-4. Participants were asked to answer: “Does this figure correctly represent part of the layout you learned from previous test?"

A memory retention score was given for each participant based on the number of correctly answered questions. For example, if they correctly answered 5 of 6 questions, they received 5 points.

![Figure 4-4. An example of pre-survey to test participants’ spatial memory retention.](image-url)
Navigation test

All participants were asked to take a navigation test with 4 trials. The order of trials is random. In the test, participants were asked to navigate to the END location they learned from session 1 as fast as they could. In each trial, they started from a new location that is different from the START location they experienced in session 1. During the test, the maze game recorded game time, travel distance, and travel trajectories of each path including positions and facing directions.

Post-experiment survey

After finishing the navigation task, all participants took a short survey describing general information and their experience with session 2. The survey consisted of three sections.

Experience on navigation task

In this section, there were 12 questions asking about some general experience on the navigation task. For these 12 questions, participants were asked to rate their agreement with 12 statements on a 1-5 scale. 5 means strongly agree, 1 means strongly disagree.

Personal information

In this section, participants were asked to answer 10 question about their personal information, including gender, age, driving experience, and so on.
**Mental rotation test**

In this section, participants took a 2D mental rotation test that is derived from Jonson’s mental rotation test (1990). Participants were presented 20 pairs of 2D images. They were asked to click the “Match” button if two images can match each other by rotating, otherwise, click the “Dismatch” button. During the test, the system recorded performing time on each question in the accuracy of milliseconds. A score of this test was given based on the number of correctly answered questions and the time to complete this test.

![Mental rotation test example](image)

**Figure 4-5.** A screenshot of mental rotation test questions from Qualtrics (answer is “Match”).

**Results of the navigation task**

In the analysis below, I used the average navigation time of four trials as the standard performance outcomes, because it could better reflect participants’ reasoning and memorizing process than navigation distance. Also, as all participants navigated at a fixed speed (moving
1000 simulated distance per second, turning 45 degrees per second) in the experiment, there is no possible confounding factor that has been criticized in previous maze tests (Morris et al., 1982).

**Declarative effects**

In this section, the effect of spatial declarative memory retention is tested and discussed. The dependent variable that was tested in this analysis is the total performance time of 4 starting locations.

An analysis was conducted to examine the total time participants took to finish the exploration task. It turns out that most of participants could finish the exploration task and figure out the possible routes in the given 10 minutes. For group 1, 17 out of 20 participants finished the exploration task in 10 minutes, and the rest of the participants (3) finished the task in 15 minutes. For group 2, 15 out of 16 participants finished the exploration task 10 minutes, and one participant finished the task in 20 minutes. Based on these times, the average time spent in the exploration task is basically equivalent between two groups. A further quantitative test such as T test is not necessary to be conducted for two reasons: first, the distribution of exploration time was not a normal distribution because most participants that finished the task faster are still asked to navigate for 10 minutes. Secondly, the exploration time is not accurately timed for participants used over 10 minutes because there is some interaction time between participants and the experimenter was also included. It is the case that they could demonstrate the same spatial knowledge.
Another set of data of the exploration task is the trajectory data that is collected by NavMaze. This data is not analyzed in this dissertation, but it is publicly available in an online repository\(^3\).

Figure 4-6 shows the performance time of 4 different starting locations, and the result is the average of all participants. The figure suggests that starting at location 2 or location 4 (check Figure 4-4) requires more time to navigate to the goal location, and this result matches the experiment layout.

![Figure 4-6. Performance time of 4 different starting locations.](image)

Figure 4-7 compares the performance time of two time intervals. The red bar represents the performance of the 7-day interval group, and the blue bar represents the performance of the 1-day interval group. From this figure, it is clear that the 7-day group took a longer time to finish the navigation task. But it is not clear whether there is a reliable difference between the 7-day group and the 1-day group on performance time because most (locations 1, 2, 4) of the blue bar is lower than the red bar, which indicates better performance.

\(^3\) [https://github.com/blayer](https://github.com/blayer)
in the range of red bar’s standard error, meaning that the value of 1 day is still in the confidence range of 7 days. After conducting a t-test between the overall performance between group 1 (Mean=85.34s, Std=36.98s) and group 2 (Mean=94.77s, Std=51.16s), the t-test result is t(34)=1.80, p=.10. This result is not reliable, meaning that the memory retention assigned (1 day vs 7 days) did not lead to different performance between the two groups after the retention interval.

![Figure 4-7. Comparison of two control groups on performance time.](image)

From the memory manipulation check, I computed the memory test scores on two conditions. For the first condition with 1-day retention interval, I found that there is no reliable difference between two visits t(34)=1.22, p=.23. This result suggests that a 1-day retention interval would not have a strong effect on participants’ spatial memory. For the second condition, with a 7-day retention interval, there is a reliable effect on scores between two visits t(34)=2.98,
p<.025), where the second visit had reliably lower scores. This significance suggests that the manipulation on the spatial memory worked.

I also tested the result of two groups in the learning session as a base line test, and it shows no reliable difference, where t(34)=0.74,p = .464. Then I tested the score differences on two groups in the testing session, and it showed a reliable difference  t(34)=2.32 p<.025). This pair of test indicates that the memory test score of two groups have no reliable difference, but their scores are reliably different after 1 day or 7 days retention manipulation. The test result can confirm that the retention manipulation worked in the empirical study.

**Procedural skills**

To test the navigation strategy that participants applied in the navigation test, I conducted a new navigation strategy scale that is adopted from Lawton’s (1996) scale on way-finding tasks. In Lawton’s way-finding navigation strategy scale, she tested a much more comprehensive navigation task that asked participants’ general experience on navigating in a town. As a result, Lawton’s scale tests orientation strategy (allocentric strategy) and route strategy (egocentric strategy) as two orthogonal scales. Different from Lawton’s scale, the scale I developed is a one-dimension scale with egocentric strategy and allocentric strategy in negative relations. This is because our experiment asked participants to take the scale with fresh impressions, but Lawton asked participants to take the scale based on assumptions and long-term experiences. Consequently, our participants would be more confirmative and negative to the unpreferred strategy by rating “not likely use it;” however, Lawton’s participants would be more neutral to the unpreferred strategy by rating “not sure.”

Based on the navigation coding strategy introduced above, I created a similar navigation score (range 1 to 5) for each participant; 5 represents applying strong allocentric strategy and 1
represents applying strong egocentric strategy. The correlation analysis is shown below, and the Cronbach’s Alpha coefficient $\alpha=.707$. The Cronbach’s Alpha indicates the internal consistency of a set of items. The Alpha coefficient ranges from 0 to 1, and a greater value indicates a higher internal consistency. Usually, it is worth to be reported when Cronbach’s Alpha coefficient is greater than 0.6.
Table 4-2. Navigation strategy scale questions with exact questions, and their scale analysis result using Cronbach's alpha analysis. ($\alpha = .707$).

<table>
<thead>
<tr>
<th>Questions on allocentric strategy</th>
<th>Correlated item – total correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>I kept track of my relative direction (e.g. left, right, front, back) to the END point.</td>
<td>0.87</td>
</tr>
<tr>
<td>Regularity or symmetry in the layout WAS NOT helpful to me.</td>
<td>0.77</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Questions on egocentric strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>When I saw a landmark, I only thought about taking-turns (e.g., turn left at the Elephant landmark).</td>
</tr>
<tr>
<td>I can only follow the landmark sequence that I remembered.</td>
</tr>
</tbody>
</table>
In the beginning of this chapter, I hypothesized that allocentric navigators can navigate faster than egocentric navigators. To prove this, I sorted the participants by their performance, and took the first quartile as strong allocentric navigator group, and took the forth quartile as strong egocentric navigator group. Naturally, the two groups were reliably different on performing time \( t(16)=39.85, p<.000 \). I also conducted a t-test on navigation score between the two groups, and the result clearly suggested that the two groups were reliably different on navigation scores \( t(16)=16.77, p<.001 \). Consequently, the significance on navigation score provides a revalidation on our navigation preference scale. In the later analysis, I mainly focused on these two groups of participants.

To examine the relation between navigation preference and navigation performing time, I conducted linear regression analysis between strategy score and performing time in the navigation test on the first quartile and the forth quartile of participants. The regression result showed a strong correlation between performing time and navigation strategy score \( (n=16, r=-.678, p<.05) \). Despite the correlation between strategy score and performing time, I also conducted t-tests to examine the effect of gender and memory retention on strategy score. But there was no reliable difference based on gender and memory retention \( t(34)=1.25, p=.23 \).

**Spatial processing ability**

Another factor I tested was the mental rotation ability that was suggested as an important issue for spatial processing (Jones et al., 1995; Malinowsky, 2001). I examined the participants’ scores \( (\text{mean}=113.19, \text{Std}=17.92) \) and finishing time \( (\text{mean}=256.75s, \text{Std}=76.22s) \). To combine the two measures, I created a transformation equation that is adopted from Johnson’s original test.
(Equation. 4-1)

\[ \text{MRT Score} = \alpha \times \text{Answer Score} - \text{Test Time} + \beta \]

(Alpha and beta are the linear coefficients to adjust the MRT score as a positive number; I use alpha = 2, and beta = 240)

The combined score had a normal distribution (mean=146.18, STD=79.24), and it matches previous results on a virtual mental rotation test (MRT) (Parsons et al., 2004). There was no reliable difference between males and females on their MRT score \( t(34)=0.98, p=.32 \). This result is the same as previous work on computer-based MRT (Parsons et al., 2004).

To test the correlation between performing time and MRT scores, I conducted a linear regression analysis on the egocentric navigation group and the allocentric navigation group. The result is on the boundary of reliable (\( n=16, r=-.453, p=.078 \)). The standardized coefficient is \( r=-.453 \), suggesting that participants’ performing time will be lower if their MRT score was higher.

Due to the unreliable linear regression on MRT result, I applied several non-linear regression approaches to gain further insights. As I found a weak reverse relation between participants’ MRT scores and performing times, it was natural to conduct an inverse regression on them. I conducted an inverse transformation of the performing time and conducted a linear regression on the MRT score. The regression result was reliable with a reliable regression fit where (\( n=16, r=.539, p<.05 \)).

To combine the two factors, I conducted a further regression analysis with MRT score and navigation score as two independent variables. The regression result suggested that these two factors could provide a reliable prediction (\( n=16, r=.799, p<.05 \)) of the navigation time with a positive coefficient on navigation score (\( \beta=0.43 \)) and a negative coefficient on the inverse of MRT score (beta = -.602).
Based on the three factors I discussed above, a simple model could be derived to roughly predict human navigation time in a familiar environment:

(Equation 4-2) \[
\text{Navigation time} = \alpha \times \text{navigation preference score} + \beta / \text{MRT score} + \theta \\
\]
(Where \(\alpha = -0.602\), \(\beta = 0.43\), \(\theta = 67.55\) in the current task)

**Results on mental rotation task**

Mental rotation task is a well-known psychology task on human visual and spatial ability. As I introduced in Chapter 1 and Chapter 2, there is a lot of work that has been done on this task. In this section, I will summarize the results from my modified mental rotation task, and the results will be reused in the next chapter for model comparison and model validation.

Figure 4-8 shows participants’ response times on 4 rotation angles. In the figure, it is clear that the response time for a 180 rotation degree is much greater than others. The response time of 90 degree and 270 degree is about the same because 270 degree clockwise is similar to rotating 90 degree counter clockwise. The fastest response time is when rotation is 0 degree, and the result of 0 rotation degree can be also seen as the baseline response time for recognizing two similar objects. The results illustrated in Figure 4-7 also match the results in Shepard and Metzler’s (1976) studies.
Figure 4-8. Effect of object rotation angles (clockwise) when number of blocks is 5. 

(n=36)^4.

In the mental rotation test, the 20 questions covered 4 rotation angles conditions, and 5 object complexity conditions. Table 4-2 shows the T-test results among 4 mental rotation angles. It is clear that all results are strongly reliable, especially the pairs between 180 degrees, 90 degrees, and 0 degree. The only pair with marginally reliable p value (p = .08) is the pair between 90 degrees and 270 degrees.

Table 4-3. T-test results between the response time of mental rotation angles (n=34).

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>90</th>
<th>180</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td></td>
<td>t = 2.47, p &lt; .025</td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>t = 3.95, p &lt; .01</td>
<td>t = 2.25, p &lt; .05</td>
<td></td>
</tr>
<tr>
<td>270</td>
<td>t = 2.28, p &lt; .05</td>
<td>t = 1.47, p = .08</td>
<td>t = 2.10, p &lt; .05</td>
</tr>
</tbody>
</table>

^4 In this dissertation, the error bar in all line charts represents standard error.
Figure 4-9 shows participants’ response times on 5 different object complexities with the number of blocks from 5 to 9. I conducted a linear regression on the result, and the regression exhibits a reliable result with regression efficiency as $r=0.978$, $p=0.04$. This result also aligns with Shepard and Metzler’s later studies on the correlating factors of mental rotation ability that response time is positively correlated with the complexity of visual object used in the task. In their study, they defined the complexity of object with the number of separate objects and projections.

![Figure 4-9](image)

Figure 4-9. Average response time of different number of blocks in the mental rotation task (n=36).

Figure 4-10 shows the response time among 5 object complexity settings; the red line represents female participants (n=15) and the blue line represents male participants (n=21). In the figure, it is clear that there is a response time gap between female participants and male participants, and it seems that the average response time of female participants is about 1 second greater than the response time of male participants. However, after conducting a T-test, there is
no reliable difference on the response time between female and male participants $t(34)=1.75$, $p=.09$. This is because the population size is relatively small and data variance is too large.

Figure 4-10. Average response time of male and female participants along different number of blocks (female n=15, male n=21).

Discussion and summary

In this work, I introduced a navigation study in a virtual environment that examines human navigation patterns and influential factors. Different from previous studies, my task was more complex and closer to realistic scenarios, as it includes both declarative knowledge of the space and procedural skills for navigation. My motivation to conduct such a complex experiment setting was that previous studies only focused on one or two aspects of humans that correlate with human navigation processes (Lawton, 1996; McNamara, 1986; Morris et al., 1982; Wang & Spelke, 2000).
Nevertheless, only a few studies have been done to compare the effects of spatial declarative knowledge retention, navigation procedural skills, and spatial retention processing abilities, and also some possible interaction effects among these factors.

In the result analysis, the total performance time is used as testing result rather than total travel distance because response time can reflect the decision making process at the decision points (intersections). Additionally, the performance time is more natural to be modeled in ACT-R because ACT-R can generate precise response time predictions.

The first question of the study was whether declarative spatial knowledge retention could influence participants’ performance in the navigation test. Specifically, I asked whether participants’ performance would get worse if their retention on the environment decreased. My initial prediction on this question was that participants should show worse performance, as what I gathered from most classical declarative experiments (Brown, 1958), because worse declarative memories on the spatial environment will increase their cognitive load and cause more recall errors during navigation. However, there was no reliably worse performance of the group with lower spatial declarative retention. The results showed that there are no reliable differences between participants with a better spatial retention (1-day interval) and those with a worse spatial retention (7-day interval). This unreliable performance difference implied that 1) declarative knowledge might not play an important role in human navigation process, or 2) some other mediators might exist between declarative spatial knowledge and navigation processes.

My second concern was around navigation strategies and the fact that it is usually considered a procedural skill. As I mentioned in the introduction, I assumed that participants only use one navigation strategy during the experiment and a navigation preference should be measured because our experiment task is relatively small and well-trained. Based on this assumption, I developed a scale to test participants’ navigation preference, and the outcome of our scale also re-validated our assumption as all questions are in the same dimension, meaning
allocentric strategy and egocentric strategy are not two independent factors. Validating the
navigation preference scale also provides us a sufficient approach to answer our second question
on how participants’ navigation preference influences their performance. According to the
analyses, our hypothesis was confirmed that participants with strong allocentric preference
showed better performance than participants with strong egocentric preference. This is because
participants who adopted allocentric representation could figure out their location and relative
direction to the goal faster when they were at a new position, then they could navigation to the
goal directly. On the other hand, participants who used an egocentric navigation strategy have no,
or less, directional sense, so they had to explore a familiar route randomly, which took more time.
Additionally, I did not find that the navigation preference score varied by spatial memory
retention or gender, meaning this preference is probably a fixed preference and an independent
factor.

In addition to the navigation preference, humans’ internal spatial processing ability
should also be taken into account for their navigation performance. The mental rotation ability
might not be the only factor of mental spatial processing but the regression result has proven it to be a sufficient standard to determine the outcomes of human navigation.

An inverse relation between mental rotation ability and performance time has been
confirmed, meaning better mental rotation ability could reliably shorten the navigation time. A possible explanation to this is that participants with better mental rotation ability could figure out the direction faster when they are facing a landmark or a path interaction from a new view angle, because this process requires participants to mentally rotate their visual memory episode to match the current view. Indirect evidence to support this explanation is that I observed participants turning their head in front of the computer to assist their mental rotation when they came to an intersection from a new direction, and this head-turning behavior was also observed during their mental rotation test. Nevertheless, the internal mechanism of mental rotation during navigation is
still unclear to us; a future study such as building a cognitive model and more data should be conducted on this issue.

A potential limitation of this study, which should also be addressed in future work, is the fact that I only manipulated participants’ memory retention in two conditions. The memory manipulation check was sufficient, but a picture recognition memory test still seems to be harder than using memory to navigate in virtual reality because a 3D virtual reality could provide much richer spatial information, such as optical cues and 3D geometry cues to help participants remember environment layouts or routes. This assumption might also explain why I cannot find any reliable effort of memory retention, because the real retention difference that participants adopted during their navigation is much smaller than what I tested in the memory manipulation checks. My future study should be focused on how to successfully manipulate participants’ spatial memory retention in the real situation, or change the experiment setting to provide a greater retention difference.

Another future direction to extend this study is to include some other standard tests on spatial processing ability, such as the Corsi block tapping task (Corsi, 1972), or visual span task (Della Sala, Gray, Baddeley, & Wilson, 1997), rather than a mental rotation test. A standard test that evaluates human visuospatial memory should be considered, because it is believed to be responsible for the temporary storage, maintenance, and manipulation of both visual and spatial information, and converting them into both visual semantic memory such as layouts, and long-term episodic memory as a visual episode of an intersection or a landmark (Jones et al., 1995). As a consequence, visuospatial memory ability could specifically influence participants’ navigation processes by altering the richness of some important visual episodes.
Chapter 5

A Modified Cognitive Model of Human Navigation Process

In Chapter 3, I introduced a navigation model using ACT-R declarative memory to implement two types of spatial mental representations, using ACT-R procedural memory to implement two different navigation strategies, and using ACT-R production rules utility theory to implement navigation strategy switching. In this chapter, I will start to compare my observation data from human participants with the predicted data from my navigation model. The initial comparison is not very satisfying, because the initial navigation model cannot accurately reflect the real navigation process in the task paradigm that I introduced in Chapter 3. But the revised model with reasonable modifications suggests the right way to improve the model fit of the data.

Criteria of judging a cognitive model

When evaluating a cognitive model, the first question would be how to judge a cognitive model, and what kinds of criteria should be followed. Different research communities have their own answers and criteria to this question. For example, in the protein science community, they compare computational protein models with observed data, and their golden measure is root mean square deviation (RMSD), which reflects how good a model can match the observed data quantitatively. The criterion for the goodness of a model is that RMSD should be less than 2 angstroms because 3 angstroms can make a reliable difference in protein level (Carugo & Pongor, 2001).

The first criterion is the originality of model. In other words, does this model exist? This is because the originality could determine how much ground work the model needs, and potential improvements to the existing model. Another important criterion is the goodness of model fitting,
because a good model fitting can provide a strong support to the hypotheses, and understanding of internal mechanisms that we care about. Dancy (2014) summarized a list of criteria to judge computational models, and they are shown in Table 5-1.

Table 5-1. The criteria to judge a cognitive model (Dancy, 2014).

<table>
<thead>
<tr>
<th>Criteria</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Does a similar model exist?</td>
<td></td>
</tr>
<tr>
<td>Does the model complete some version of the task?</td>
<td></td>
</tr>
<tr>
<td>Can the model complete the same task (including software) that was used to collect experimental data?</td>
<td></td>
</tr>
<tr>
<td>How long does it take for the model to run?</td>
<td></td>
</tr>
<tr>
<td>What level of behavior does the model describe?</td>
<td></td>
</tr>
<tr>
<td>Does the model simulate processes involved or is it solely focused on output?</td>
<td></td>
</tr>
<tr>
<td>How well do model predictions match experimental data? And what data are matched to the experiment</td>
<td></td>
</tr>
<tr>
<td>If there have been previous versions of the experiment, can the model give you predictions that have not been found in previous versions of the experiment, but that can be recorded in the future?</td>
<td></td>
</tr>
</tbody>
</table>

In the cognitive modeling community, there are two widely accepted measures to determine the goodness of model fitting: correlation coefficient (r or $r^2$ value) and root mean square deviation (RMSD).

The correlation coefficient is a qualitative measure of some type of correlation between the observed data and prediction of the model. In the other words, the correlation coefficient represents how much the prediction model can explain the observed data. The value of correlation coefficient is usually known as r value, and it ranges from -1 to 1. A positive r value represents positive correlation between observed data and model prediction, accordingly, negative value represents negative correlation. Usually, we only use a positive r value in model fitting, and a
high r value means a better fit. A widely accepted truth in the community is that a model fitting with $r > 0.9$ would be considered a good fit, but it also depends on the originality and complexity of the model.

Root mean square deviation (RMSD) is a quantitative measure that represents the accuracy of a model prediction. Mathematically, RMSD tells the average difference (residual) between the model prediction and observed data, and it can range from 0, meaning 100% fit, to infinity, meaning a huge gap. In the community, we expect RMSD to be as small as possible, but it seems there is no agreement on the criterion of RMSD. One criterion from an engineering perspective suggests that a good RMSD value should be less than 10% of the mean value of observed data. Alternatively, another criterion from a statistical perspective suggests that an RMSD value less than one standard error of observed data can be considered as good. In my opinion, there should be a fixed criterion, and it should depend on the measures (mean and standard error) of observed data. If 10% of mean is much greater than standard error, or standard error is much greater than 10% of mean, we should take the smaller value of them as the criterion of goodness. If 10% of mean is almost the same as one standard error, either criterion would be acceptable.

In the rest of this chapter, I review some comparisons between model generated data and empirical data from Chapter 3. I use the criteria discussed above to judge my data fittings, and make some improvements based on my data fitting result.

**Comparing navigation model with empirical data**

This section covers the comparison between the predictions of NavModel and the empirical data. It also includes the introduction of experiment settings, and experiment platform.
VIPER

In this experiment, I used VIPER (Kaulakis et al., 2012), a virtual implementation of plural environmental representations. This platform was initially created for simulations of large-scale social-cognitive agents (Zhao, Kaulakis, et al., 2011; Zhao, Kaulakis, Morgan, Hiam, & Ritter, 2012; Zhao et al., 2015). For this dissertation work, I modified VIPER to simulate the spatial environment that I created in the NavMaze experiment platform. This VIPER-based environment has the same layout and landmarks as NavMaze but all of this spatial information is coded in plain text format rather than 3D views. I chose to make this simplification due to two limitations of ACT-R, and all other cognitive architectures.

1) ACT-R cannot process 3D information in its visual module, because it only has a visual representation of a 2D screen.

2) ACT-R cannot process complex visual information such as textures, images, and visual dynamics because its visual module only supports a limited number of object formats. Most visual information has been hardcoded into plain text.

However, VIPER as a text-based virtual environment provided me a flexible spatial platform to interact with the NavModel agent since VIPER can provide spatial information in the plain text format that can be easily encoded into ACT-R chunks. VIPER was initially created to be a lightweight spatial world to model the evolution and behavior of complex social networks where you would like to simulate how agents move in a 2D world to generate and exercise networks. So it provides a lot of flexibility to modify its layout and objects. VIPER also has a comprehensive direction system that provides spatial information from both egocentric and allocentric formats.

VIPER was created in C based on Naked MUD. The simulation handles complex computations on the server side, but uses additional Python scripting. This Python scripting allows for an extremely flexible implementation of ideas, including persistent objects (modifiable objects that
retain their changes even across runs). Further, this scripting model allows for the simulation itself to call on outside programs to launch and create events within the environment. An example of this feature would be a specific testing area that would call for a Java program to populate itself with scripted agents. These additional agents provide additional layers of complexity by providing many more interactions for the intelligent agents. Agents connect and interact in VIPER using the Telnet protocol (RFC 854). Languages such as Java and LISP provide robust Telnet libraries. VIPER runs asynchronously; the speed and frequency of communication is determined by each agent. VIPER is designed so that variations in performance originate from the agents participating in the environment, as opposed to being a function of the environment. VIPER does not and cannot record the internal state of agents in it because it has no access to this information. VIPER, however, does record temporal and spatial behaviors of agents, as well as agent interactions, including strings passed between agents. Logs are saved in a CSV file for use in tools like ORA (Carley, Pfeffer, Reminga, Storrick, & Columbus, 2013).

**Model settings**

I used a MacBook with a 1.4GHz i5 processor and 4G memory to run VIPER and a NavModel agent. I used Java 1.6, Python 2.7.10, and ACT-R 6.0 in my operating system. The port number for connecting VIPER is set as 8000. In ACT-R, the learning parameter :bll is set as default 0.2, declarative memory noise :ans is disabled at the beginning, and the utility reward parameter is set as default as 0.1.

The experiment layout was rebuilt in VIPER as a 17*17 map with a total 289 rooms, and 97 rooms actually connected. Each room matches a 1000*1000 simulated distance in NavMaze. All rooms were named by their location in the maze. For example, the upper left room shown in Figure 5-1 is named as 0507, representing it is in row 5 and column 7. Consequently, a NavModel
agent can quickly localize itself when entering a room because the room name will be given in a text dialogue format to the agent. For example, if the agent is entering room 0507, a text string will be sent to the agent with the content of “enter room: 0507.”

![Figure 5-1. NavModel experiment layout implemented in VIPER.](image)

For each round of experiments, I ran the model to start from locations 1, 2, 3, and 4 (locations 1-4 are indicated in Figure 5-2.). For each starting location, there was at least one shortest path to the end location. Between these 4 shortest paths, a portion of these paths overlap with each other. Consequently, if the starting location follows a fixed sequence, for example, 1-2-3-4, the response time of location 2 should be shorter than running the route from location 2 as the first run because the run of location 2 could reuse memory retention of the run of location 1. To avoid this learning effect, the sequence of 4 starting locations was generated randomly. Due to limited time and computation resources, I repeated the experiment 10 times, because each round of experiment took about 30-40 minutes for the model to run.
Figure 5-2. Experiment layout with starting locations and end location.

Comparison of model prediction and empirical result

Figure 5-3 shows the total experiment time that participants and the NavModel used to finish tasks 1 to 4. In the figure, the empirical data is presented in the blue solid line, NavModel prediction with $:ans=0$ is presented in the red dashed line, and NavModel prediction with $:ans=0.5$ is presented in the green dotted line.
Figure 5-3. Comparison between empirical data and NavModel1.0 generated data ($ans=0$, $r=0.419$, $RMSD=57.09$; $ans=0.5$, $r=0.449$, $RMSD=19.28$).

The correlation between empirical data and prediction with $ans=0$ is ($r=0.49$, $r^2=0.17$, $RMSD=57.09$). This prediction is a poor fitting as both measures are very far from the satisfactory criteria. To get a better fit, I increased the $ans$ value that represents declarative memory noise, because all participants received a memory intervention in the experiment, and their declarative memory should have already been decayed with some disruptions or noise. The correlation between empirical data and prediction with $ans=0.5$ is ($r=0.49$, $r^2=0.20$, $RMSD=19.28$). Based on these measures, this seems to be another poor fit with no reliable improvement. I stopped trying to change ACT-R parameters to explore the best fit because the model prediction is quite far from the empirical data, meaning that there are definitely some structural mistakes in the NavModel.
Discussion

Based on the human data introduced in Chapter 4, I found that the initial navigation was not accurate in two aspects:

First, the initial navigation model used room number as a reference to locate itself, and this approach is simplified from real process. Based on the comparison shown in Figure 5-3, it is clear that the curve of human data and the curve of model prediction has a gap on performance time, because the initial model simplified the self-orientation and self-localization processes. However, as the gap between prediction and human data is relatively large, it means that the cost associated with self-orientation is an important mental process, and it should be included to improve model prediction fit. Usually, the explanation provided for this process involves a variant of mental rotation, because it has been reported repeatedly (Gunzelmann & Anderson, 2004; R Shepard & Hurwitz, 1984) that a variety of orientation tasks claimed that mental rotation is involved in transforming a representation of the information in one reference frame to match the orientation in another reference.

Second, the initial model implemented a map navigation strategy based on a map-based spatial representation, and used this strategy as a secondary supplement to route-based strategy. However, based on the results introduced in Chapter 4, I found two facts that are contradictory to my initial assumptions on navigation strategy.

First, most of the participants (30/38) claimed that they only applied one navigation strategy during the task, and they did not switch to a secondary strategy when they failed to recall the routes. In the group of single strategy participants, there are a large proportion (21/30) of participants are classified to be map-based strategy adopters, and a small proportion (9/30) of participants are classified as route-based strategy adopters.
Second, when single strategy adopter failed to recall spatial information, they tended not to use any secondary strategy, except a random search or visual search. This result indicated why some participants took a much longer time to finish the task because they failed to recall route information, and started to randomly explore the maze until they found some location that was retrievable.

**Mental rotation model**

Based on the discussion above, I will first improve the NavModel by adding a reliable mental rotation model. This section will introduce the details of the mental rotation model and the model fitting process with empirical data that was collected from my mental rotation task.

**Mental rotation model description**

The mental rotation model used the ACT-R GUI Interface to re-implement the experiment interface that I used in my mental rotation test. As the ACT-R GUI Interface is a native implementation of computer screen, objects can be easily accessed by the visual module of ACT-R. The only difficult part of this implementation is that I had to hardcode all objects, including visual stimuli objects and report buttons, because the GUI Interface was implemented in native Lisp, and can only accept input setting in Lisp script format.

Figure 5-4 shows a high-level illustration of the mental rotation model. This model contains 5 DM chunks and 24 rules in ACT-R. The model first looks at the left image and loads it into imaginal memory as a temporal storage. Then, the model uses imaginal memory and the Mental Rotation function to rotate the left image with a run-time delay (delay to stop the entire ACT-R timing). In the Mental Rotation function, it can calculate new coordinates of all objects
and wait for proper latency based on the rotation angles. After finishing all mental rotation in imaginal memory, the model moves its visual attention to the right image, and matches each object between the rotated image and the image in visual attention. If some objects fail to match or no more objects can be matched, a decision can be made and the model starts to call the manual module of ACT-R to perform a mouse action to select “Match” or “Mismatch,” then click the “Next” button.

Figure 5-4. The brief structure of mental rotation model v1.

Figure 5-5 shows a detailed illustration of the ACT-R model, and there are some noticeable details I want to introduce.

1) Before the model encodes all image information, including positions of blocks and colors into image-left chunk. This chunk is designed to hold all the blocks’ information, and store it in imaginal buffer. Before starting the matching process, the model also moves its attention to the right image, and does the same encoding process with image-right chunk.
2) The model uses declarative memory to record the correct rotation angle in the \textit{rotation-angle} chunk. The \textit{rotation-angle} chunk is loaded into declarative memory when the model starts. If values of this chunk is nil, the model starts a heuristic matching first to find the correct rotation angle. When a rotation angle is matched, the model updates \textit{rotation-angle} and uses it for later rotation.

3) The mental rotation process in imaginal memory actually rotates single blocks each time rather than rotating the entire image. This is due to the limitation of the ACT-R visual modules, but it also matches the empirical fact that human attention is limited and cannot process a complex image at one time. The block information is stored in \textit{color-block-left} chunk, and it is passed from the visual module imaginal to the module for mental rotation process.
Figure 5-5. The logic flowchart of the mental rotation model.
Mental rotation function

Equation 5-1 shows a mental rotation function that computes new coordinates and simulates latency based on rotation angles. This function is implemented as an extension of the imaginal memory of ACT-R in Lisp, and it uses the hook of the imaginal buffer to pass the rotation result and simulated latency to ACT-R. The imaginal-action hook mechanism is a function that enables ACT-R to call external functions or a thread with a self-defined delay time.

The imaginal module of ACT-R is a system that is designed to store and manipulate context relevant to current tasks. The module consists of two buffers: 1) an imaginal buffer that typically maintains additional context for the current task, such as visual input or auditory input; 2) an imaginal-action buffer that is usually used for users to extend the capability of the imaginal module. The imaginal-action buffer does not perform any action itself, but it provides a flexible hook for modelers to perform any actions on imaginal buffer. The imaginal-action hook mechanism is a function that enables ACT-R to call external functions or a thread with a self-defined delay time. Imaginal-action buffer is essentially same as a call to ‘!eval!’ in the production system, because they all send a call to the functions of the production rules. The only difference is that the imaginal-action buffer provides a parameter that the modelers can modify simulation delay that is used by mental rotation function to implement latencies of mental rotation processes.

In the imaginal module, there is an :imaginal-delay parameter that controls how long it takes to complete actions in the imaginal module. This parameter enables the mental rotation function to control the delay time in the mental rotation process. Equation 5-1 illustrates how delay time is determined in the mental rotation function. In the equation, :base-time presents delay time for recording information into imaginary module, and it is set as 200ms as the suggested value by ACT-R; degree represents rotation difference between left image and right
image; \( l \) represents the slope between delay time and rotation degree, and this value varies by the complexity of tasks.

Equation 5-1

\[
T = base-time + l*degree
\]

In the mental rotation function, there are two parameters modelers that can change, which are :\( base-time \) and :\( l \). The default setting for both parameters are 200ms and 9ms/degree (Lyon et al., 2008).

Comparing mental rotation model with empirical data

As I introduced at the beginning of this chapter, validating a model with empirical data is the most crucial part of a cognitive modeling process. In ACT-R there are at least 110 parameters (based on ACT-R 6.0) I can adjust, so the model fitting process is more likely to be an exploration process, and this exploration process itself is also worth reporting.

Default settings

In the experiment testing mental rotation results, I used all the ACT-R default settings for the declarative module and the procedural module in ACT-R. The subsymbolic function of declarative memory is turned off because there is no expected learning effect involved in this experiment. The partial matching function of procedural memory is turned on, because there might be a procedural learning effect in this multiple-round experiment. For imaginal action, I set the action latency time as 200ms to present the time that used for block matching action imaginal buffer. This latency time was similar to Halverson, Gunzelmann, Moore, and Van Dogen’s (2010) mental orientation model in ACT-R. The rotation slope was set as 9ms/degree, and it is
similar to the slope found in a similar psychology research (Bothell-Fox & Shepard, 1988). I ran the ACT-R model in simulation mode along 20 different task settings (see Appendix B) for 30 times, and there was no variation in the result because this model didn’t use any sub-symbolic functions in ACT-R. The result and comparison are shown in Figure 5-6 and Figure 5-7.

Figure 5-6. Model fitting on rotations angles with default setting (r=.81, RMSD=3.61).
Figure 5-7. Model fitting on the complexity of rotation object with default setting (r=.97, RMSD=3.48).

Figure 5-6 shows observed reaction time per rotation angles with a blue line, and model prediction with a red line. The model fitting is basically well (r=.81, RMSD=3.61), but it still cannot meet the criterion that is usually used in our cognitive modeling community (r should be better than 0.9). Also, the RMSD is not very satisfactory because there is a big gap between the observed result and the model predictions.

Figure 5-7 shows observed reaction time per number of blocks with a blue line, and model prediction with a red line. The model fits the observed results very well, as r=.97 (p>0.01). However, RMSD (=3.48s) is also not very satisfactory, with a total reaction time of 12s. Based on Bothell-Fox and Shepard’s (1988) suggestion, the slope between the reaction time and rotation degree varied from 8ms/ degree to 30ms/degree, and it is closely related to the complexity of
rotation objects. So, the value of slope would be the first parameter to manipulate in model fitting.

**Best fits**

To improve fitting results of the mental rotation model, I increased the slope value from 9ms/degree because a large value of slope could result in a high reaction time, which could lower the gap between the observed result and model predictions. Also, the smaller gap could result in a small RMSD, which means a better model fit. In my model fitting process, I increased the slope value from 9ms/degree to 22ms/degree, and the model fitting result is shown in the following two figures.

![Figure 5-8. Model fitting on rotations angles with best parameter setting (r=.823, RMSD=1.03).](image-url)
Figure 5-9. Model fitting on the complexity of rotation object with best parameter setting (r=.94,RMSD=1.10).

Figure 5-8 shows the new model fitting result on four rotation angle settings. Comparing it with Figure 5-6, I find that the results of the new model improved in both qualitative level and quantitative level, where the r value has a slight improvement from .801 to .823, and RMSD has a satisfied improvement from 3.61 to a more acceptable value at 1.03.

In Figure 5-9, the data fitting looks better because the RMSD value decreases into an acceptable range from 3.48 to 1.10. This means that this mental rotation model can explain the complexity of rotation object very well in both qualitative level (r=.94) and quantitative level (RMSD=1.10). It can also support that the mechanism I used to separate rotation objects and to match individual sub-objects in this model can correctly represent the internal process of human cognition for this mental rotation task.

Based on the result comparison above, it seems that using a slope parameter as 23ms/degree can generate a much better fit. However, this slope parameter is not consistent with
previous psychology empirical data. Shepard and Metzler (1988) conducted a comprehensive study to examine the range of mental rotation speed and related factors. They summarized their result and all previous experiment results in the table below.

Table 5-2. A summary of response time per degree in existing mental rotation experiment.

<table>
<thead>
<tr>
<th>Tasks and objects</th>
<th>Shepard and Metzler’s (1988)</th>
<th>All previous experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>One stimulus</td>
<td>2D 2.1 ms/degree</td>
<td>6.4 ms/degree</td>
</tr>
<tr>
<td>(compare one object)</td>
<td>3D 2.9 ms/degree</td>
<td>N/A</td>
</tr>
<tr>
<td>Two stimuli</td>
<td>2D 6.5 ms/degree</td>
<td>14.3 ms/degree</td>
</tr>
<tr>
<td>(compare two objects)</td>
<td>3D 7.75 ms/degree</td>
<td>16.5 ms/degree</td>
</tr>
</tbody>
</table>

In Table 5-2, I found that 23 ms/degree is much slower than all previous empirical results as the highest slope parameter is 16.5 ms/degree, and it was derived from the result of 3D objects. In my mental rotation experiment, I used two stimuli and 2D objects, so the slope parameter of best model fitting is about twice much as previous empirical results.

As the best model fitting is contradictory to empirical data, I cannot accept it as a successful and useful model. To improve the mental rotation model, the only possible way is to extend or adjust the mechanisms of mental rotation model based on the comparison between empirical data and model predictions. In the next section, I introduce a new mechanism based on comparison and observation of participants.
**Adjusted mental rotation model**

The above section introduced two the model fitting processes based on changing parameters in mental rotation function, and it kept all ACT-R parameters as default. I noticed that the model prediction failed to fit the empirical data in several aspects: first, there is a constant gap between empirical data and model prediction. This means that the model missed an internal process, and this process is independent from all known variables such as rotation angles, number of blocks, and object dimensionality. Additionally, as the gap between empirical result and model prediction is relatively large (around 3 seconds), it might be a repeated process in every block matching. Second, in Figure 5-4, the qualitative measure between empirical result and model prediction is not acceptable as the $r = .82$. This is because the model prediction on 270 degrees was far from the empirical data. In my observation of the empirical data, the mental cost of rotating 180 degrees should be higher than 270 degrees, meaning the current mechanisms on modeling rotation over 180 should be modified in ACT-R.

Based on my discussion above, I adjusted the mental rotation model, and the basic structure is shown Figure 5-10.
Figure 5-10. The workflow of adjusted mental rotation model.
Comparing Figure 5-5 and Figure 5-10, there are two noticeable differences that correspond to two reasons for the unsatisfactory model fittings discussed above.

The first modification I included in the model is that I re-implemented the mental rotation process in the imaginal buffer, and applied a visual search mechanism in it. As the model prediction is much less than the observed data, it means that the mental rotation model version 1.0 oversimplified some mental process that reduced simulation time. Just and Carpenter (1976) suggested that eye movement between objects and fixations on an image is highly correlated with the time of mental rotation, and eye movement should be the main explanation for Shepard and Metzler’s mental rotation experiment, rather than manipulating a mental image in human cognition. Based on their argument, one possible explanation might be that the 1.0 model simplified the visual search process by moving the entire left image information into the imaginal buffer. To correct this oversimplification, I re-implemented the mental rotation process from rotating the entire image to rotating single blocks (colored squares in my experiment) of the image. As it shows in Figure 5-11, the new implementation moves attention to a single block and rotation of this block in the imaginal buffer; meanwhile, it moves attention to the block of the right image and starts to match. When a block matched, the model searches for a new block of the left image, and starts the matching loop again. In Figure 5-11, it shows the process of matching single objects and attention moving.

The second modification is adding a rotation direction slot in the declarative memory. In the 1.0 model, the default rotation direction is clockwise because the mental rotation function starts to explore rotation angle from the clockwise direction. When a correct rotation angle is found, the 1.0 model directly records this clockwise angle into declarative memory; as a result, the rotation angle would be recorded as 0, 90, 180, and 270. However, the empirical study shows that humans exhibit the same ability to mentally rotate objects between clockwise direction and counter-clockwise direction (R. Shepard & Hurwitz, 1984; R. Shepard & Metzler, 1971), which
means that 270 of clockwise should be equal to 90 of counter-clockwise. This also explains why the 1.0 model prediction has a reliable difference from the observed data at 270 degree (see Figure 5-6 and Figure 5-8). With the second modification on the rotation direction slot, the 2.0 model is able to differ rotation directions, and produce a more accurate prediction.

![Diagram](image)

**Figure 5-11.** The brief structure of the adjusted mental rotation model (version 2.0).

**Comparing adjusted mental rotation model with the empirical data**

After proposing the adjusted mental rotation model that processes separate objects, I implement this model with 5 DM chunks and 26 rules in ACT-R. I compared the predictions of the mental rotation model 2.0 with the empirical data.
**Model settings**

In the experiment of this adjusted model, I still used most of the default settings in ACT-R. The subsymbolic function of declarative memory was turned off because there was no expected learning effect involved in this experiment. The partial matching function of procedural memory was turned on, because there might be a procedural learning effect in this multiple-round experiment. In mental rotation function, the base rotation time was set as default value as 200ms, the rotation slope value was set as 9ms/degree at beginning.

One more noticeable setting that is different from the 1.0 model is that the 2.0 model replaces the standard visual module with EMMA-an ACT-R visual module extension with eye movement and visual encoding (Salvucci, 2001). The EMMA model typically focuses on a more rigorous account of eye movement and attention shifting and their interaction in the main cognitive processor of ACT-R. There are two main components that EMMA includes: eye movement and visual encoding. The eye movement component describes the temporal and spatial characteristics of eye movements, because they are responsible for attention shifting. The visual encoding component describes the effects of frequency and foveal eccentricity when encoding visual objects as internal representations. With these two components and their interactions, EMMA could generate behavioral predictions with more precision, from high-level attention shifting to low-level eye movement.

EMMA was initially implemented as an individual model of ACT-R, and it was eventually merged into the ACT-R standard distribution since ACT-R 6. To enable this function, I have replaced the default visual module file with the EMMA.lisp file and set a value for the :vis-obj-freq parameter that controls visual attention shifting frequency. In this experiment, I set :vis-obj-freq at a default value of 0.01. The experiment procedural was same as the 1.0 model’s; I
ran 20 experiment settings that matched with human studies, and repeated the experiment 20 times.

Model fitting

In the experiment, I also adjusted the slope value to explore the best fitting result. When \( l \) is equal to 11ms/degree, I found the prediction result to be close to the best fitting. Based on the criterion shown in Table 5-1, 11ms/degree of slope is between Shepard and Metzler’s survey result (6.5ms/degree) and all experiment average result (12.3ms/degree) for two stimuli+ 2D experiment settings. This means that the slope parameter of model 2.0 is in the acceptable range and it matches previous empirical data. The result is shown in the Figure 5-12 and Figure 5-13 below.

Figure 5-12. Adjust mental rotation model fit on rotation angles (\( r=.93, \text{RMSD}=1.09 \)).
Figure 5-13. Adjust mental rotation model fitting on number of blocks

\[(r=.95, \text{RMSD}=1.21)\].

Figure 5-12 shows prediction results of model 2.0 on number of blocks, and I found that \(r=.93\) and \(\text{RMSD}=1.09\). Based on the model fitting criterion that was discussed at the beginning of this chapter, both \(r\) value and \(\text{RMSD}\) are good enough to be reported. In Figure 5-13, the prediction of model 2.0 on rotation angles also fits well with empirical data, with \(r=.95\) and \(\text{RMSD}=1.21\). The \(\text{RMSD}\) is slightly off the mark as it is slightly greater than 10% of average response time on rotation angles (average =9.65), but all prediction values are still in the range of standard error. So, the fitting result is still acceptable in most criterion systems.

I have introduced three different versions of mental rotation models and their prediction results. As every model has its advantages and disadvantages in some aspects, I have to compare three models and further discuss which model would be appropriate for the navigation data. This comparison and discussion is addressed in the next section.
Discussion

As I discussed at the beginning of this chapter, $r$ value (or $r$ squared) and RMSD are considered a golden standard in cognitive modeling. Most publications in the cognitive modeling community try to convince readers by showing graphs of empirical data along with model prediction. The model fit is assumed to be convincing if two graphs are similar. From a statistical point of view, a good fit is assumed to be a high $r$ squared value and lower RMSD. In practice, a good model fit is usually based on an exploratory process of parameter testing. For example, in the mental rotation model 1.0, I adjusted the slope parameter from 9ms/degree to 25 ms/degree to get the best model fit.

However, these traditional model fit methods or criterion seems to not be applicable and convincing because there are some hidden issues with these methods. One issue is that this exploratory process is not always a feasible option because complex models can take substantial time to run a single set of parameters, and testing all combinations seems to be very expensive. Another issue is that most models with multiple parameters have a considerable freedom to be adjusted, such as adding a new production or changing a set of parameters. With these two hidden issues, researchers (Taatgen & Rijn, 2010) argue that the traditional fitting process and criterion cannot always be followed. The alignment of previous theories or empirical data, and the understanding of model’s internal cognitive mechanism should also be seriously considered, when judging a model.

Based on the discussion above, I summarize three high-level criteria to judge a cognitive model, especially when cognitive modelers are facing multiple models with different parameter settings and cognitive mechanisms.
1) Of course, the conventional measures including r or r squared value and RMSD should be
decent. They probably are not the best fits, but still in the acceptable ranges that were
introduced at the beginning of this chapter.

2) All parameters should be in the acceptable ranges or consistent with previous psychology
research data or existing cognitive models. For most cognitive architectures such as
ACT-R and Soar, their system parameters are well-defined and well-restricted with
previous research data. But it does not mean that models can adjust all parameters as
much as they want to fit the model, because some values of parameters presents a
particular extreme case. For example, when modeling a vocabulary learning experiment
that was conducted in the lab, it is not appropriate to set an :ans parameter that presents
mental noise to be 2.0 because a value of :ans over 1.0 is designed for heavy external
manipulating, and existing learning model usually set this parameter below 0.5. For some
parameters that are presented by module extension or added functions, they should be
carefully adjusted to match empirical data that the module extension is built upon.

3) The internal cognitive mechanism should match with existing theories or reflect a new
understanding of a theory. When we create a cognitive model, the essential goal is not to
fit a particular data graph, but to explain the phenomenon we are interested in. Although
most of the phenomenon cannot be fully explained by existing theories, it is still very
possible to link the cognitive model with existing theories.

Table 5-3 compares the three models with all measures, advantages, and disadvantages.
All good measures are highlighted in bold. Based on the high-level criteria above, I decide to
select Model 2.0 as the best model explaining the empirical data gather from my mental rotation
experiment, because this model is the only model out of the three with all measures and
parameters in the acceptable range. Additionally, it implements a cognitive mechanism that uses
object separation, visual search, and attention shifting. More importantly, this new cognitive mechanism with proper data fitting might provide us with a deeper understanding of how humans mentally rotate objects, providing a promising answer to the long-term debate on the explanation of mental rotation effect.

Table 5-3 Comparison of three versions of the mental rotation model (bold represents “best” fit).

<table>
<thead>
<tr>
<th></th>
<th>Model 1.0 default settings</th>
<th>Model 1.0 best settings</th>
<th>Model 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>r value (angle/blocks)</td>
<td>0.81/0.97</td>
<td>0.82/0.94</td>
<td>0.95/0.93</td>
</tr>
<tr>
<td>RMSD (angle/blocks)</td>
<td>3.61/3.48</td>
<td><strong>1.03/1.10</strong></td>
<td>1.21/1.09</td>
</tr>
<tr>
<td>Base rotation time</td>
<td>200 ms</td>
<td>200 ms</td>
<td>200 ms</td>
</tr>
<tr>
<td>Rotation slope</td>
<td><strong>9ms/deg</strong></td>
<td>23.5ms/deg</td>
<td>11ms/deg</td>
</tr>
<tr>
<td>Advantages</td>
<td>All ACT-R default parameter values</td>
<td>Best fits</td>
<td>Acceptable fittings, acceptable parameters, improvement implementation in imaginal chunks</td>
</tr>
<tr>
<td>Disadvantages</td>
<td>bad RMSD, simple implementation in imaginal chunks</td>
<td>Out of range rotation slope, seems to be over fitting</td>
<td>Not best fitting</td>
</tr>
</tbody>
</table>

As I introduced at the beginning of this section, one the original finders of mental rotation ability, Shepard and Metzler (1971), explained this phenomenon by a “mental imaginary” facility that helps humans to temporally store and manipulate the visual objects to rotate from any axis. The response time is only related to the rotation angles, and the correlation is perfectly aligned. In
contrast, Carpenter and Just (1973) criticized the theory of “mental imaginary” and explained the perfect correlation between response time and rotation angles by distances of eye movements. In this section, I implemented the mental rotation process in two approaches: the mental imaginary approach (model 1.0) and the hybrid approach with mental imaginary, object separation, and attention shifting (one behavior of eye movement). The results suggest that the hybrid model achieved a reliably better fitting with empirical data, which means that mental rotation process does involve mental imaginary as Shepard explained, but the mental imaginary might only be responsible for handling separated key element of the rotated objects. Another important cognitive process in mental rotation is visual search of all key elements; for example, the colored blocks in my experiment are a kind of simplified key element. Attention shifting should also be taken into account of mental rotation as Carpenter and Just argued, however, eye movement is just a sub-part of mental rotation because it is only related to attention shifting between images and matching blocks.

**Integrating the mental rotation model in navigation model**

Figure 5-14 shows a modified model of NavModel; it is named NavModel 2 to differentiate it from the initial NavModel. This model includes 4 DM chunks and 45 rules in ACT-R. According to the discussion of the first section of this chapter, I conducted several modifications reflecting my further understanding of the empirical results.
Figure 5-14. An illustration of NavModel’s structure.

First, the rewarding mechanisms of strategy shifting were removed because the empirical survey reveals that participants do not claim any strategy shifting in the experiment. Also, the utility rewarding mechanisms fail to exhibit strategy shifting during experiment. Based on my analysis of the ACT-R logs, I cannot find any experiment trial that has applied route-based strategy because agents are placed at off-route starting locations, and they have to be applied to a couple rounds of map-based strategy to navigate back to a familiar route. At the point, map-based strategy already becomes a dominant strategy because it receives a lot of positive rewards. As a result, agents apply map-based strategy during the entire experiment, and the total simulation time can be highly correlated with distance between the start location and the goal location because the ACT-R agents allow the application of the map-based strategy.

Second, to replace the rewarding mechanism, I applied a fixed preference with route-based strategy, goal directed strategy, and random strategy. This preference correlates with
participants’ survey results that most of them only applied a route-based strategy, but they also kept a direction reference to the goal location. As a result, when agents were placed in an out-route start location, they were still able to navigate toward the direction of the goal or navigate back to a route that they remember. Nevertheless, the route-based strategy is not a pure sequential strategy that was introduced in Chapter 1 because it also interacts with the self-orientation module that requires a local configurational view. The local configurational view does not represent the layout of the entire environment but a small and standard location unit such as a room. In this dissertation, all configurational representations refer to the map view of the entire environment, and all configurational strategies use the map view of the entire environment. Nevertheless, I still use the term of “route-based strategy” in this model introduction to differentiate it from map-based strategy.

Third, to keep a direction reference to the goal location, the model has to orient itself in the maze. Based on survey that was introduced in Chapter 4, most of participants admitted that they kept a direction reference to the goal, meaning they could always tell which direction the goals. However, keeping this reference required participants to locate themselves. In NavModel 1.0, VIPER numbered all rooms by their location, and the NavModel agent could directly orient itself by computing the room number. However, room number with location encoded seems to be an oversimplified approach that the prediction, causing the NavModel 1.0 to be very far from empirical data. In NavModel 2.0, I modified the VIPER system to make it output egocentric spatial information such as “outlet: back, left, right”. This egocentric visual creates a more realistic spatial environment, and it requires a more complex cognitive process that is believed to be related to mental rotation behavior. Also, I created a self-orientation module based on the mental rotation model 2.0 that was introduced in the section above. In this self-orientation module, the model converts egocentric spatial information from VIPER into view-info chunk, and compares it with map-based memory to determine orientation. Once the orientation is determined,
it will be stored in goal-direction slot of goal chunk. Every time the agent makes an action, such as turning left, the goal-direction slot will be updated.

**Comparing extended navigation model with empirical data**

**Model settings**

All settings were the same as the experiment of NavModel 1.0, except the output of VIPER has been changed from the allocentric view such as “outlet—north, south” to the egocentric view such as “outlet—back, left.” Also, all room names have been changed to remove all location information.

For each round of experiment, I ran the model to start from locations 1, 2, 3, 4. To avoid the learning effect, the sequence of the 4 starting locations was generated randomly.
Comparisons between model prediction and empirical result

Figure 5-15. Model fitting of total performance time of 4 starting location (:ans=0, r=.811, RMSD=35.00; :ans=0.5,r=.889, RMSD=22.78).

In the Figure 5-15, the empirical data is presented with a blue solid line, NavModel prediction with :ans=0 is presented with a red dashed line, and NavModel predictions with :ans=0.5 is presented with a green dotted line. The correlation between the empirical data and predictions with :ans=0 is (r=.811, r^2=.66, RMSD =35.00). Compared to the result of NavModel 1.0, the result with :ans=0 has a reliable improvement, but the r value and RMSD are still not very satisfactory. The correlation between the empirical data and predictions with :ans=0.5 is (r=.889, r^2=.79, RMSD=22.78). The correlation coefficient for the second prediction is close to the acceptable boundary but the RMSD measure is still far from satisfactory.

During exploration for the best fitting, I find that the r value and RMSD would be improving when increasing :ans value. But I think the result with :ans=0.5 is close to the best
fitting so far because the standard error is already much greater than the empirical data when :math:`\text{ans}=0.5`. If I continue to increase :math:`\text{ans}` too much, some experiment runs could end up with memory failure, and that would reliably increase standard error.

**Summary**

In this chapter, I presented a navigation model in ACT-R named NavModel, and some discussions on model improvement and parameter exploration were also included. At the beginning of the chapter, I compared the prediction of NavModel 1.0 with the empirical data, and the fitting result is fair to poor. Based on participants’ surveys and previous research, I extended the NavModel 1.0 with a sub-model of mental rotation process, and a new mechanism of navigation strategy selection.

While building the mental rotation model, I created two versions of the model to reflect the aspects of existing theories on mental rotation test. The prediction fitting with empirical data reveals that the second version of the mental rotation model with mental imaginary and attention shifting involved generates better model fittings, and all parameters are in the reasonable range.

The new strategy selection mechanism applied a fixed preference and direction reference in the declarative memory. It also integrated the mental rotation model as a self-orientation module in the NavModel 2.0. The prediction of NavModel 2.0 fits empirical data fairly well, especially when declarative memory noise is enabled and set as 0.5.

The criteria that was discussed at the beginning of this chapter indicated that the second version of the mental rotation model was a good fit and it is worthy to be reported and reused. The NavModel 2.0 had a slightly less fit but it is still useful as it is the first model comprehensive navigation model in ACT-R that integrates spatial representations, self-orientation, and navigation strategy. Also, the NavModel 2.0 interacts with a more complex environment, which
is very close to realistic scenarios, providing a very promising platform for future modeling research on navigation behavior. Based on current version, the NavModel can still be improved in several ways, such as adding a manual (hand) module to interact with VIPER, but it still can be judged as a good model, and should be very useful to share with the community.

During the model development and parameter exploring process, I also found some interesting phenomenon that might be able to provide some new insights to existing research questions.

First, the comparison between the two versions of mental rotation models reveals that the traditional explanation of mental rotation process might not be accurate because the vision-based mental imaginary theory could not fully explain the empirical data. By combining mental imagery and attention shifting together, the model is able to generate a more accurate prediction. This might reveal that attention shifting during mental rotation also plays a very important role. I cannot confirm that this new mechanism is the definitive explanation of mental rotation behavior because of a lack of new empirical data for verification, and some known limitations of ACT-R. But the new mechanism of mental rotation still provides a very promising explanation to be verified in the future.

The second interesting finding is that the NavModel provides a new explanation on human navigation preference selection and related factors. The explanation that generated from NavModel 2.0 is contradictory to my initial hypothesis that humans conduct navigation strategy shifting during navigation. Actually, the good fit of NavModel 2.0 confirms Tar’s (2005) finding that humans apply fixed navigation preference on route-based strategy. Nevertheless, the model also confirms that the mental rotation ability can play an important role in the navigation process as it helps human to conduct self-orientation and locate themselves. This important finding of mental rotation ability is also aligned with the analysis of Chapter 4, in which it shows a high correlation between total navigation time and participants’ mental rotation scores.
Chapter 6

Conclusion

In this chapter, I review this dissertation with some general insights of each study and model that were presented. I also summarize the contributions of this work from both empirical and modeling perspectives. Some potential ways to extend this work and some known limitations of this work are also discussed at the end of this chapter.

Empirical navigation study

The empirical navigation study focuses on examining the influential factors of navigation process. Based on the previous literature and the simulation results of the NavModel, the study is designed to examine three main factors that are spatial memory retention, navigation preference scores, and mental rotation ability.

To complete such a study, I implemented the online navigation experiment platform NavMaze to build classical psychology maze study in a virtual reality. I created a maze study in NavMaze with 13 landmarks and 4 starting locations. For memory retention, the study was designed with a learning phase and a testing phase, and participants’ spatial memory retention was manipulated by assigning different time intervals between two experiment phases. For navigation preference, a novel navigation preference scale was designed to evaluate participants’ navigation preference score. For mental rotation ability, all participants were asked to take an online mental rotation test that simplifies classical mental rotation studies.

The results showed that spatial memory retention exhibits no reliable influence on navigation process because there is no reliable performance difference between participants with
a 1-day interval and participants with a 7-day interval. Nevertheless, the preference score and mental rotation score exhibited high correlation with navigation performance, as the correlation coefficient \( r = 0.799 \) is high under the standard of behavior study.

**Mental rotation model**

As the analysis of empirical navigation study indicates that participants’ navigation time is highly correlated with their mental rotation test score, this dissertation focuses on a deeper understanding of the cognitive process of mental rotation. In this dissertation, a mental rotation model is described that completes a 2D mental rotation test, and the model is able to fit the empirical data very well.

In the subsymbolic level, the mental rotation model implements an external function in Lisp that simulates response time of mental rotation process. The external function is also hooked up with the imaginal module of ACT-R, which is used as a mental image buffer in the mental rotation model. In the symbolic level, the mental rotation model describes how declarative memory, visual module, imaginal module, and manual module interact over the 2D mental rotation process. I have implemented two versions of the mental rotation mental with different symbolic mechanisms. The comparison between the empirical data and the predictions of two models provides a novel explanation of the mental rotation task that humans attend and rotate separate key features of the figure, rather than memorize and rotate the entire figure.

This model is one of the few models in ACT-R that involves declarative components, visual/manual components, and imaginal components. Also, it is the only running ACT-R model that implements a mental rotation task in both symbolic and subsymbolic levels. To fit the empirical data, I had a set of explorative experiments on subsymbolic parameters. The value of
parameters could be reusable in similar tasks because this is the first running ACT-R model and it fits empirical data.

**Navigation model**

This dissertation primarily focuses on the explanation of the mental process of navigation and influential factors by describing a cognitive model, NavModel, in the ACT-R architecture. The NavModel is implemented to complete a comprehensive navigation task that is similar to the most popular maze navigation tasks in psychology. Due to the limitation of the ACT-R architecture, my colleagues and I first implemented VIPER, an experiment platform for navigation tasks. This platform provides a text-based and highly editable environment, and a telnet protocol to communicate with cognitive agents. On the agent side, we created a communication protocol in Lisp that can be easily integrated into the ACT-R agent. As a result, the VIPER platform enables researchers to create a navigation task in any layout, and enables their ACT-R model to directly “see” and “move” in VIPER.

According to related literature on navigation studies that were introduced in Chapter 1, NavModel implemented two spatial memory representations: a map-based representation and a route-based representation. Based on these two representations, NavModel also implements three navigation strategies: a map-based strategy, a route following strategy, and a random search strategy. To explain human navigation preferences, I implemented two versions of NavModel, where NavModel 1.0 applied a procedural learning strategy, and NavModel 2.0 applied a fixed preference with the route strategy as the primary navigation strategy. The comparison between model predictions and empirical data suggests that NavModel 2.0 is better than NavModel 1.0.

The explanation that was generated from NavModel 2.0 contradicts my initial hypothesis that humans conduct navigation shifting during navigation. Actually, the good fit of NavModel
2.0 confirms Tar’s (2008) finding that humans apply fixed navigation preference on route-based strategy. Nevertheless, the model also confirms that mental rotation ability plays an important role in the navigation process, as it helps human to conduct self-orientation and locate themselves. This important finding of mental rotation ability is also aligned with the analysis of Chapter 4, in which it shows a high correlation between total navigation time and participants’ mental rotation scores.

**Contributions of this work**

I summarize at least ten contributions of this work, and the paragraphs below discuss these contributions in four aspects.

**An online platform to navigation experiment**

NavMaze is a light-weight navigation experiment platform based on HTML and JavaScript. I developed this platform for my dissertation empirical studies because nearly all virtual reality experiments that have been used for previous studies are not reusable or editable. NavMaze provides a highly compatible and reusable platform for all similar navigation studies because it is implemented in pure HTML and JavaScript, which are runnable in all Internet browsers. The output can be plain text format or .csv format, and it could be read by most popular data analysis software, such as Excel, Matlab, or R. In addition, NavMaze is also highly editable, in which researchers could edit map layout, output format, and experiment scenarios by changing a couple of parameters and a few lines of JavaScript code.

Based on my knowledge and research, NavMaze is the only working experimental navigation system online. Right now, it has drawn some attention from the online learning and
education community. Researchers from TellLab, an online psychology experiment platform that was developed by the Department of Psychology at Harvard University, are integrating NavMaze into their system. The source code of NavMaze is also available at Github for developers and researchers to download\(^5\).

**A study examining influential factors on navigation process**

This dissertation introduces an empirical study based on NavMaze and Qualtrics. This study is one of the few studies on complex navigation behavior and influential factors. In addition, this study is more reusable than most traditional navigation studies because researchers can quickly replicate the experiment on TellLab by only changing a few parameters, or can easily extend the experiment on GitHub and Qualtrics.

The experiment results also shows the factors that influence navigation performance that deeper our understanding of human navigation processes. The results suggest that spatial memory retention does not have influence on navigation performance, which differs from my initial hypotheses. But it still confirms that mental rotation ability and navigation preference score is positively correlated with navigation performance, and the correlation is fairly strong.

**A cognitive model of mental rotation task**

This dissertation describes a mental rotation model in ACT-R by adding a mental rotation function that is hooked up with the imaginal module of ACT-R. As this mental rotation function is verified by empirical data, it could be reusable for later models that involve the same mental

\(^5\) [https://github.com/blayer/3DMaze](https://github.com/blayer/3DMaze)
rotation process. In that case, such a supplement function would enhance the competence of ARC-R architecture, and also provide a strong supplement to the unified theory of cognition.

Based on the mental rotation function, a complete mental rotation model in ACT-R is also implemented. This model is the first ACT-R model that completes a standard mental rotation task, and the prediction of this model is convincing.

During the model testing process, a more interesting contribution of this model is that the traditional explanation of the mental rotation process might not be accurate because the vision-based mental imaginary theory could not full explain empirical data. By combining mental imagery and attention shifting together, the model is able to generate a more precise prediction. This might reveal that attention shifting during mental rotation also plays a very important role. I cannot confirm that this new mechanism is definitely the accurate explanation of mental rotation behavior, because of the lack of extra empirical data for verification, and some known limitations of ACT-R. But the new mechanism of mental rotation still provides a new and deeper perspective to understand the cognitive process of mental rotation task.

A cognitive model to complete a complex navigation task

This dissertation describes a cognitive model in ACT-R named NavModel, and it is one of a few ACT-R models that can complete complex navigation tasks in a virtual environment. The contribution of this model can be summarized from two perspectives.

From an engineering perspective, this model provides a complete environment platform for future studies of modeling navigation tasks in ACT-R. VIPER provides a highly editable environment; theoretically, it can be modified for any layout or scenario because all layouts, objects, and actions are text-based and can be easily understood and edited. The communication protocol between VIPER and core ACT-R is a generic connector that uses a text-based format; in
this case, researchers can apply this connector to any telnet communication between core ACT-R and any environment or software using telnet protocol.

From a theoretical perspective, this model is one of a few ACT-R models that can complete a complex navigation task, and it is the only model that involves mental rotation process. During the modeling process, it also provides a novel explanation on human navigation strategy preference and adoption. The good fit of NavModel 2.0 is consistent with Tar’s (2005) finding that humans apply fixed navigation preference on route-based strategy. Additionally, the model also suggests that mental rotation ability plays an important role in navigation process.

**Limitations and future directions of this work**

This work still has some limitations, and I can only pick up a list of major limitations for discussion. In addition, I also provide some future directions to overcome these limitations and to extend this work in some promising ways.

**Empirical study**

As I introduced in Chapter 3, the memory manipulation test is designed to verify the effect of manipulating spatial memory. The result shows a reliable difference between two groups of participants with different memory manipulation. However, there is no reliable difference on any performance measures between the two groups of participants. This result might be influenced by two limitations in the experiment that are 1) the effectiveness of memory test, and 2) the population size of empirical study.

In the memory test, the questions are in the true/false format asking about landmarks and high-level layout. However, the context of these questions might not be very relevant to the
spatial knowledge that participants used in the real experiment because the context of questions focused on 2-dimensional allocentric spatial knowledge rather than 3-dimensional egocentric knowledge. The irrelevant memory test used in the empirical study might not be accurate enough to reflect the real spatial knowledge that guides participants during navigation, and it should be a potential work for the future.

The limited number of participants might be another reason for the unsatisfying result. In my study, I was only able to recruit 38 participants due to limited time and the relatively heavy workload for participants. As I explained in Chapter 3, the experiment required participants to take two sessions within two days, so the drop rate (52 signed up, but 14 dropped) of this experiment is higher than other experiments. Also, the entire experiment requires at least 2 hours for participants, which makes them less motivated to sign up, and more difficult for participants to attend. To recruit as many participants as possible, the most efficient way is to provide them a high extra credit (2%) at the end of the semester. But I could only recruit from a limited number of classes due to instructors’ policies and I only had two weeks to run the experiment.

Another limitation of the empirical study is the mental rotation test. In my test, participants only had a tutorial section with two sample questions; this might be not enough for them to fully understand the mental rotation test. As a result, a procedural learning effect might be involved in the empirical result and it has not been considered in the mental rotation model. In a future study, there are two potential ways to solve this problem: 1) I have to re-run the mental rotation test with a longer tutorial section to avoid learning effect compounding with the effect of blocks and rotation angles, 2) or I can also modify the mental rotation model by enabling learning. It might help to improve model fit. Gunzelmann and Anderson’s model (2004) on the procedural learning effect in a self-orientation task generates a very accurate prediction, and including their model might be a good start.
Additionally, one limitation is the data analysis of the empirical study only takes the total performance time as the main indicator. However, it is hard to differentiate between a delay caused by visual search and a delay caused by decision-making. Based on the current data that collected during the empirical study, a possible future analysis that can be examined to differentiate visual search and decision-making would be the log data of NavMaze in the testing session. In the log data, it contains a set of participants’ location coordinates and facing directions, by which, I can identify the items that participants were looking when there was a pause.

Another possible future work is to use some different measures in the testing session rather than the performance time. For example, Gillner and Marlot (1998) used layout sketching to study spatial knowledge acquisition in a virtual maze. They introduced a valid evaluation of configuration knowledge by evaluating the number of errors and sketching quality ranks in the entire controlled group. In the future work, a secondary evaluation of spatial representation memory such as sketching should be included because the retention of spatial representation can effectively influence strategy selection and shifting (Bibby, 1989). More specifically, the representation learned in this case map could force use of map-based strategy. Being tested on routes, might give rise to a route-based strategy strongly. Bibby showed this even true after learning.

A potential limitation of this study, which should also be addressed in the future work, is the fact that I only manipulated participants’ memory retention in two conditions. Our memory manipulation check was sufficient, but a picture recognition memory test still seems to be harder than using memory to navigate in the virtual reality because a 3D virtual reality could provide much richer spatial information, such as optical cues and 3D geometry cues to help participants remember environment layout or routes. This assumption might also explain why we cannot find any reliable effect of memory retention because the real retention difference that participants
adopted during their navigation is much smaller than I tested in the memory manipulation checks. My future study should be focused on how to successfully manipulate participants’ spatial memory retention in the real situation, or change the experiment setting to provide a greater retention difference.

Another future direction to extend this study is to include some other standard tests on spatial processing ability, such as the Corsi block tapping task (Corsi, 1972), or visual span task (Della Sala et al., 1997), rather than a mental rotation test. A standard test that evaluates human visuospatial memory should be considered, because it is believed to be responsible for the temporary storage, maintenance, and manipulation of both visual and spatial information, and convert them into both visual semantic memory such as layouts, and long-term episodic memory as a visual episode of an intersection or a landmark (Jones et al., 1995). As a consequence, visuospatial memory ability could specifically influence participants’ navigation process by altering the richness of some important visual episodes.

The last future direction of the empirical study is to conduct a further analysis on the participants’ trajectory data that is collected during the exploration task. Such trajectory data and facing direction data would be able to reflect some other factors that determine navigation strategy preferences. Actually, analyzing the facing directions is still too weak because it is hard to determine which objects participants are looking at by analyzing facing directions. A more promising way is to use eye tracking as a supplementary device to NavMaze, so researchers are able to analyze the key object that participants used during exploration and navigation.

**Mental rotation model**

One of the limitations of the mental rotation model is that the model was not communicating with the original experiment software. In the cognitive modeling community, one
of the criteria to judge a cognitive model is whether the model can produce same behavior in the same experiment settings. The ideal experiment settings would have the model communicate with the same software, and produce the same result. However, I conducted my empirical study on Qualtrics as a webpage-based experiment rather than software running on my local machine. Due to the limitation of ACT-R and its supplement software, it was hard to connect the mental rotation model and Internet Explorer. To make this model more solid, a future direction of this dissertation would be to re-implement my mental rotation task as a Java or Lisp application that communicates with ACT-R model through a local telnet port.

Another limitation was that I didn’t explore further into visual attention shifting and eye movements. As I introduced above, the mental rotation model 2.0 utilizes EMMA, a supplementary module of ACT-R that simulate eye movements while attention shifting. But I just applied the default parameters of EMMA without exploring the parameters for a better fit. As noted in the instruction of EMMA (Salvucci, 2001), the default parameters were summarized from empirical results that were gathered from a single stimulus, and it might be different from our mental rotation task with multiple visual stimuli. I believe that adjusting parameters in EMMA would improve model fitting and, more importantly, it would provide a deeper understanding of eye movement patterns in mental rotation tasks.

**VIPER and NavModel communication**

The major limitation of NavModel is the communication format between NavModel and VIPER. As I explained above, NavModel receives text-based information from VIPER, and information of the environment, such as objects, room numbers, and directions are directly updated in the goal buffer rather than the visual buffer that humans usually do. This is due to the limitation of the visual module in ACT-R, because the visual buffer and visual-location buffer in
the visual module of ACT-R can only access information that is already defined in the model.

Another limitation of the ACT-R visual module is that the visual-location buffer is only able to perceive and store 2-dimensional information; more specifically, it can only perceive left, right, up, and down. As a result, if I use the visual module to receive information from the environment VIPER, the depth information will be lost. Actually, the limitation of ACT-R’s visual module has been discussed at great lengths (Gunzelmann & Lyon, 2007) in the past decade, so extending the visual module in multiple ways would be a popular direction for the community.
Appendix A

Instruction letter for the navigation task

Task Instruction (Session 1)

In this task, you will be instructed to complete a list of navigation tasks in a 3D game. Please read the following instructions that includes basic steps of the experiment. The experimenter will also be guiding you during the experiment.

Tutorial

The participants will play with the game using the keyboard in front of them and are asked to finish 4 basic tutorial tasks, including moving, turning, and using assisting information. This tutorial takes 5 mins in total, and the experimenter will confirm with every participant that they fully understand how to operate the software.

Learning Task

Participants have to learn three pre-defined routes with a starting location and a goal location repeatedly for 5 mins. All routes will be highlighted in the 3D environment with some yellow arrows. Then, participants need to re-navigate the routes without highlighted arrows.

Task Instruction (Session 2)

Learning Task continued

When coming back after a retention interval, participants will take a free exploring task in the environment with a virtual map layout shown on the upper right corner of their screen. Participants from group 2 will be randomly assigned to a start location, and asked to goal location repeatedly for 5mins or at least 5 successful navigations. When navigating, they will be able to check their current location and the goal location in the virtual map.

Navigation Task
After finished the second part of learning task, all participants from both group are asked to navigate from a randomly given location to the pre-defined goal location.

**Detour task**

After finishing the navigation section, all participants will finish the same navigation task with all routes broken out by a virtual wall. Then, participants are asked to navigate from a randomly assigned location to the goal location again.
# Appendix B

**Mental rotation test images**

<table>
<thead>
<tr>
<th>Question 1</th>
<th>Question 2</th>
<th>Question 3</th>
<th>Question 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 blocks</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>6 blocks</td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
</tr>
<tr>
<td></td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
<tr>
<td>7 blocks</td>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
<td><img src="image15.png" alt="Image" /></td>
</tr>
<tr>
<td></td>
<td><img src="image16.png" alt="Image" /></td>
<td><img src="image17.png" alt="Image" /></td>
<td><img src="image18.png" alt="Image" /></td>
</tr>
<tr>
<td>8 blocks</td>
<td><img src="image19.png" alt="Image" /></td>
<td><img src="image20.png" alt="Image" /></td>
<td><img src="image21.png" alt="Image" /></td>
</tr>
<tr>
<td></td>
<td><img src="image22.png" alt="Image" /></td>
<td><img src="image23.png" alt="Image" /></td>
<td><img src="image24.png" alt="Image" /></td>
</tr>
<tr>
<td>9 blocks</td>
<td><img src="image25.png" alt="Image" /></td>
<td><img src="image26.png" alt="Image" /></td>
<td><img src="image27.png" alt="Image" /></td>
</tr>
</tbody>
</table>
Appendix C

List of memory retention test questions

Questions of layout memory:

1. Does the figure below reflect the correct map layout you have learnt?

![Diagram 1](image1)

2. Does the figure below reflect the correct map layout you have learnt?

![Diagram 2](image2)

3. Does the figure below reflect the correct map layout you have learnt?

![Diagram 3](image3)
Questions of layout memory:

4. Is the animal below on the route from the start location to the end location?

5. Is the animal below on the route from the start location to the end location?
6. Is the animal below on the route from the start location to the end location?
Appendix D
Navigation preference scale

We want to know your experience in the task you just finished. Please rate your agreement with the following statements on a 1-5 scale. 5 means strongly agree, 1 means strongly disagree, if you are unsure, enter a 2, 3, or 4 depending on whether you are totally neutral or leaning toward agreement or disagreement.

1) I could easily visualize the maze layout.

2) When I saw a landmark, I could tell its position (e.g. center, bottom left, top right of the maze) in the layout.

3) I kept track of my relative direction (e.g. left, right, front, back) to the END location.

4) I kept track of my relative direction (e.g. left, right, front, back) to my START location.

5) When I saw a landmark, I only thought about taking-turn (e.g., turn left at the Elephant icon).

6) When I was positioned at a new START, landmarks that were NOT on the two previously learned routes COULD NOT help me to the END location.

7) Regularity or symmetry in the layout WAS not helpful to me.
References


Dancy, C. L. (2014). *Why the change of heart? Understanding the interactions between physiology, affect, and cognition and their effects on decision-making.* (Ph.D Unpublished Thesis), Penn State, University Park, PA.


Stevens, C. A. (2014). *Are you sure the library is that way? Metacognitive monitoring of spatial judgements*. (Ph.D), Penn State University, USA.


Changkun Zhao

EDUCATION
The Pennsylvania State University 2016
Ph.D., College of Information Sciences and Technology
Shanghai Jiao Tong University 2007
B.S., College of Electrical and Information Engineering

PUBLICATIONS
Journal Articles, peer reviewed

Conference Publications, peer reviewed

AWARDS
Recommended reads (best paper nominee) of BRIMS
2012
Travel Award to HCIC 2015
2015