

# **Decision in Space:**

## **Reinforcement Learning Models Predict Individual Differences in and Neural Correlates of Strategy Adoption during Spatial Navigation**

Submitted by: ir. Dian Anggraini, M.Sc.

Date of Submission: 15 May 2018

Dissertation der Graduate School of Systemic Neurosciences  
der Ludwig-Maximilians-Universität München

Supervisor: Dr. Klaus Wunderlich

Co-supervisor: Prof. Stefan Glasauer



First Reviewer: Dr. Klaus Wunderlich

Second Reviewer: Prof. Stefan Glasauer

External Reviewer: Prof. Shu-Chen-Li

Date of Oral Examination: 20<sup>th</sup> September 2018

## **Thesis Advisory Committee (TAC)**

Dr. Klaus Wunderlich  
Prof. Stefan Glasauer  
Dr. Virginia Flanagan

## **Research Lab**

*Emmy-Noether-Gruppe Wunderlich*  
Department of Psychology  
Martiusstr. 4, 80803 Munich, Germany

*Center for Sensorimotor Research*  
German Center for Vertigo and Balance Disorders  
Klinikum der Universität München  
Feodor-Lynen-Str. 19, 81377 Munich, Germany

## **Research Funding**

*Emmy-Noether Grant*  
This research was primarily funded by the Deutsche Forschungsgemeinschaft [DFG] through the Emmy Noether Grant (WU 758/1-1)

*German Center for Vertigo and Balance Disorders*  
This research was also supported by Bundesministerium für Bildung und Forschung (BMBF) Grant (01 eO 0901).

*Graduate School of Systemic Neurosciences*  
The GSN-LMU provided support for attending conferences, summer schools, as well as various courses and scientific meetings that are crucial for the success of the doctoral work.



# Dedication

To Kartika Utami Nurhayati and Alya Kusuma Nurjannah, Stellka and Karinka Trusinová, Tinka Dériková, Aksekainaat and Ujala Nazar Jamil, Leia Olivia Wunderlich, and Vika Neumann. So that this doctoral thesis becomes an inspiration for the next generation of women in Neurosciences.



# Acknowledgement

## ***Looking Back***

There were moments when I thought this day would never come. I prayed for its quick delivery, crossed days off my calendars, counted hours, minutes and seconds, and now that it's here, I am both grateful and puzzled. I am especially grateful for the people I met throughout this scientific journey. I am also puzzled because by finally submitting this doctoral thesis, I might have to leave some friends who inspire me and researchers who have been my mentors. As you read through this thesis, it is my humble hope that it becomes clear that this work is the result of excellent collaboration between different researchers and tremendous help from friends, family, and all participants who willingly took part in my experiments.

## ***What a Scientific Adventure!***

First and foremost, I owe my deepest gratitude to my supervisor, Dr. Klaus Wunderlich, for trusting me and giving me the opportunity to develop exceptional skills through this PhD work. He has been a transformative presence in all aspects of my education. Despite my apparent stubbornness and ambition, I really appreciate his gentle encouragement and the fact that his office was always open for me. I am so thankful for his guidance, patience, advice, enthusiasm, as well as enormous scientific insights and expertise. Above all, Klaus has given me a new perspective on science and on life that I would otherwise not have.

Being an engineer myself, it was extremely encouraging to collaborate with another engineer. In that respect, I would like to thank my co-supervisor, Prof. Stefan Glasauer, for submerging me in the topics of spatial navigation, for the interesting discussions, and

his enormous enthusiasm. Thank you very much for the unremitting support and confidence in finding my own path to get the work done.

Undoubtedly, my ‘job’ as a doctoral researcher came with its excitements as well as frustrations. For this, I am forever grateful to Dr. Virginia Flanagin. Thank you for all the insightful conversations, valuable comments, encouragement, and above all for trusting my ability especially when I was most in doubt.

Moreover, it has been an honor and a great pleasure to be part of the Graduate School of Systemic Neurosciences. Thanks to the GSN, I have had the chance to participate in various soft skill courses, summer schools, and interesting conferences. A bucket of appreciation is due to Lena, Stefanie, Renate, Catherine, Raluca, Birgit, Julia, Nadine, Alex, and of course Benedikt. Thank you for always being there in my corner, supporting me along the way.

### ***I Can Never Do This on My Own***

Next, I would like to thank those I had the pleasure to collaborate and who joined me during a significant part of my scientific journey.

To Kathrin Kistorz for providing me with unlimited access to her coffee supply as well as her attentive pair of ears. To Nisha Dalal who never fails to share insightful comments and her love for Bollywood movies. To Theresa Raiser who completely understands my frustrations when it comes to working with *the machine*. To my friends and colleagues at the Forschungshaus, Psychology Department, AMGEN, and the GSN (Cora, Ke, Hilary, Luisa, Eli, Liliana, Ot Prat, Natalia K, Beth L, Judita, Christopher, Thommi, Peter, Alicia, Simone, Isi, Josh, Alex, Lina, Angela, Michaela, Fernando, James, Steffi, Matthias, and Rainer) – thank you, not only for your professionalism and advices, but also for the jokes and friendships. You all made a time consuming work bearable (sometimes even enjoyable).

To my colleagues at the Institute for Stroke and Dementia, especially Miguel and Mathias. To Dr. Viliam Rapčan for providing me with countless scholarly inputs. To those I met at conferences and became great friends. To the people who volunteered their time and were willingly lying inside the MRI machine for hours so that I could get



the data I need for this thesis. It has been an honor to get to know and work with this superlative group of individuals.

### ***The Walk of Life***

On my way called life, I meet lots of people of whom some in particular deserve special attention. Without these supporters and personal cheerleaders, I would most likely not have reached the point where I am today.

I would like to thank five individuals who have been my role models throughout this entire journey: Carsten and Claire Neumann, Teresa Vania Tjahja, Ariane Boehm, as well as Dr. Vanessa Johnen. Knowing you and calling you my friends are truly the highlights of my PhD work. It is truly an honor to share life changing experience as well as important milestones in life. Thank you for always encouraging me to be the best version of myself.

During the final stage of my PhD, I had the pleasure of getting to know an amazing individual who then became the source of my drive and motivation. Westley Villalobos is a kind, decent, and unfailingly generous person. I thank him for all the jokes and conversations that we shared. Above all, I owe him my sincere gratitude for his constant reminder that ‘every end is a new beginning’.

Throughout this doctoral work, there were times when I had doubts. There was a moment when I lost connection with Neuroscience. There was even an instance when I hit rock bottom and thought that my work did not matter. Fortunately, I met a group of outstanding individuals who picked me up, helped me regain my confidence, showed me the beauty of science, and reminded me of the very reason why I chose this scientific path. For this, I would like to thank my friends at 15x4 Munich: Adeel Yawar Jamil, Aleksei Zinovev, Evgeny Labzin, Slava Edelev, and Viktoria Korzhova. It has been a great honor and privilege to work with all of you. Each and every one of you constantly amazes me with all that you do. Thank you for all the intellectual conversations that we shared, constructive criticisms that we threw at each other, and a phenomenal commitment to communicate science to a wider audience.

Since the beginning of my doctoral research, the Rapčanová (Pan Viliam a Pani Edita, Editka a Katka) and the Spuchláková have given me unconditional love and support. They have been by my side during these last few years; making sure I was relaxed enough in the summer time and warm enough during the cold winter months. In addition, they have been giving me the motivation to succeed. For this, I am forever grateful.

I would also like to take a moment to thank the Hutches: Mike, annie, KT, and of course my Mom and Dad (Mary-Beth and Lou). It is a simple fact that I wouldn't have come to this point without their love and supports. The one-year time I spent with them in the USA opens a lot of doors for me. I am especially grateful that they taught me not only to have pride in my work, but also to believe in it and to cherish every moment of it. I hope someday I can change someone's life as bravely and selflessly as they have changed mine.

Certainly, I would like to thank my mother and my two little sisters. Of all people they invested the most in me, giving me a head start on the highway called life. Since my early age, my mother is my ultimate inspiration in life. She never failed to teach me to have pride and confidence in my work. She never gave me any idea that I couldn't do whatever I wanted to do or be whomever I wanted to be. She filled my childhood with love and fun, books and music, as well as knowledge and endless hours of foreign language courses (I am forever grateful). I thank her deeply for her unparalleled support and prayers. As she guides me through these incredible 30 years, I don't know if she ever realized that the person I most want to be is her. To my two little sister (Dina and Dini) who never fail to be my number one fans, thank you for all the stories, all the dreams, and the hopes that we share.

Last but certainly not least, thanks to you, my dear reader. By now you already read at least four pages of my thesis. Thank You.. (You are almost there.)





## **Space: the final frontier.**

These are the voyages of the starship Enterprise.

Its continuing mission: to explore strange new worlds.

To seek out new life and new civilizations.

To boldly go where no one has gone before

*(Star Trek the Next Generation)*



# Summary

For both humans and animals, the ability to orient oneself in an environment is essential for everyday functioning. This ability to navigate is affected by individual differences, such as age, gender, and cognitive strategies adopted for orientation. When it comes to strategy adoption in spatial navigation, the brain can either track a well-known route or plan a novel path using cognitive representation of the environment. These two strategies are known as route-based and map-based navigation. A similar dual strategy approach exists for value-based decision making. One strategy, the model-free choice, is a repetition of previously successful behavior. In contrast, the model-based choice uses an internal representation of task structure to assess which future decision will lead to the most highly valued outcomes. The computational mechanisms of these two strategies in value-based decision making have been well explained using the Reinforcement Learning (RL) algorithms.

To investigate the interplay between neural processes involved in spatial navigation and value-based decision making, we combined a Virtual Reality (VR) wayfinding paradigm, fMRI, and computational models based on RL algorithms. The core of our wayfinding task was a grid world where participants navigate in search of specific target objects. Participants were first allowed to freely explore the environment. They then proceeded through three experimental phases. These three phases encouraged the use of either route-based or map-based navigation without limiting the participants to either navigational strategy. We then modeled participants' navigation strategies, as adapted to reach target objects, using either model-free, model-based, or a combination of the model-free and model-based RL algorithms. The key internal variables from RL modeling were then used as parametric regressors to elucidate the computational mechanism of strategy adoption during navigation.

This cumulative thesis consists of two manuscripts that studies computational mechanism and neural correlates of spatial navigation as explained by value-based decision making strategies. In the first manuscript, we tested the hypothesis that different navigational strategies relate to different neural computational mechanisms as is the case for making model-free and model-based choices. We found that participants' choice behavior during the wayfinding task could be well explained by the RL algorithms. We also demonstrated that the BOLD signal in certain brain regions correlated well with the RL key internal variables, i.e. the model-free or model-based value signals. BOLD signal in the left ventromedial prefrontal cortex (vmPFC), retrosplenial complex (RSC), and caudate nucleus pertained to model-free value signals. In contrast, BOLD signal in the right parahippocampal gyrus extending to the medial temporal lobe (MTL), precuneus, and left RSC pertained to model-based value signals. In addition, we also showed that the BOLD signals in the left vmPFC and right parahippocampal gyrus were particularly well explained by the RL value signals in participants with propensity to model-based over the model-free choices.

In the second manuscript, we introduced another element to our study, namely age. It has been extensively reported that navigation ability is among the most severely affected cognitive faculties in both normal aging and dementia. Thus, we conducted a similar experiment, as the one presented in the first manuscript, on two groups of participants: older adults between the age of 60 to 75, and younger adults between the age of 23 to 35. The younger participants performed the wayfinding task on a 24" monitor in a lab setting. The older participants performed the same task while undergoing functional Magnetic Resonance Imaging (fMRI). In our task, we found that while younger participants used a mixture of route and map-based navigation across trials, the older participants would exhibit strong propensity towards route-based navigation. We also showed that the traversed routes in both groups of participants could be well explained by either the model-free or model-based reinforcement learning (RL) models. Significant correlations between navigation indices and key parameters from the RL models confirmed that RL models account for individuals' variability in strategy adoption during navigation. With regard to neuroimaging results, correlation between BOLD activity and the model-free value signals provided evidence that the model-free RL algorithm explain how the aging brain computationally prefers to rely more on route-based navigation.



In brief, in spite of the use of RL and fMRI navigation paradigm for the past decades, there has been little integration of the two approaches, especially in human. Thus, this work leads to improved knowledge about the utility of RL algorithms to identify both the computational mechanism and neural correlates of strategy adoption in spatial navigation and wayfinding. Lastly, the studies described in this thesis are important to advancing the field of both spatial navigation and value-based decision making. This is because, both studies show the potential use of the RL models to improve the modeling of various cognitive processes involved in spatial orientation and/or cognition.



# Contents

Dedication .....	v
Acknowledgement .....	vii
Summary .....	xv
Contents .....	xix
<b>1 General Introduction .....</b>	<b>33</b>
1.1 The Concept of Wayfinding and Spatial Navigation .....	36
1.1.1 Reference Frames in Wayfinding and Spatial Navigation .....	38
1.1.2 Organization and Properties of Spatial Knowledge .....	39
1.2 Strategy Adoption in Spatial Navigation and Wayfinding .....	42
1.2.1 General Concept of Route-based and Map-based Navigation .....	42
1.2.2 Age-related Changes in Navigation Strategies .....	44
1.3 Assessing Wayfinding Strategies Using Virtual Environment .....	46
1.3.1 Navigation in Virtual vs. Real Environment .....	46
1.3.2 Assessment Paradigms in Virtual Environment .....	50
1.4 Neural Basis of Wayfinding Strategies .....	57
1.4.1 Medial Temporal Lobe (MTL) .....	59
1.4.2 The Striatal Complex .....	62
1.4.3 The Retrosplenial Complex (RSC) .....	63
1.5 Value-based Decision Making .....	67
1.5.1 Cognitive Processes in Value-based Decision Making .....	67
1.5.2 Valuation Based on Model-Free Choice .....	68
1.5.3 Valuation Based on Model-Based Choice .....	69

1.6	Reinforcement Learning as a Computational Framework to Represent Choice Behavior .....	72
1.6.1	Model-free Reinforcement Learning .....	73
1.6.2	Model-based Reinforcement Learning .....	78
1.6.3	Action Selection.....	79
1.7	Neural Correlates of Reinforcement Learning during Value-based Decision Making81	
1.7.1	Prefrontal Cortex.....	82
1.7.2	Basal Ganglia.....	87
1.8	Aim of the Thesis .....	89
<b>2</b>	<b>Manuscript 1: Neural Signatures of Reinforcement Learning Correlates with Strategy Adoption during Spatial Navigation .....</b>	<b>93</b>
2.1	Abstract.....	94
2.2	Introduction .....	95
2.3	Results .....	99
2.3.1	Strategy Adoption during Wayfinding Task.....	99
2.3.2	RL Model Fits to Navigation Data .....	101
2.3.3	Neural Signatures of Model-Free and Model-Based Choices.....	102
2.4	Discussion.....	106
2.5	Methods .....	111
2.5.1	Participants .....	111
2.5.2	Navigation Indices.....	111
2.5.3	Reinforcement learning model .....	113
2.5.4	fMRI Data Analysis.....	115
2.5.5	Region of Interest Analysis .....	116
2.6	Figures.....	118
2.7	References.....	123
2.8	Supplementary Information.....	131
2.8.1	Supplementary Results and Discussion.....	131

---

2.8.2	Supplementary Methods .....	133
2.8.3	Supplementary Figures .....	138
2.8.4	Supplementary Tables .....	142
2.8.5	Supplementary References.....	146
<b>3</b>	<b>Manuscript 2: Of Reward, Age, and Space: Reinforcement Learning Algorithms Explain Age-related Changes in Spatial Navigation .....</b>	<b>149</b>
3.1	Abstract .....	150
3.2	Introduction.....	151
3.3	Results.....	154
3.3.1	Spatial Navigation Strategies in Two Groups of Participants .....	154
3.3.2	RL Model Fits to Navigation Data.....	156
3.3.3	BOLD Signals Pertaining to Value of Chosen Action .....	158
3.4	Discussion .....	159
3.5	Methods.....	163
3.5.1	Participants.....	163
3.5.2	Procedure .....	163
3.5.3	Navigation Indices .....	164
3.5.4	Reinforcement Learning Model.....	165
3.5.5	fMRI Data Analysis .....	167
3.6	Figures .....	169
3.7	References .....	175
3.8	Supplementary Information .....	180
3.8.1	Supplementary Results and Discussion .....	180
3.8.2	Supplementary Methods .....	181
3.8.3	Supplementary Figures .....	186
3.8.4	Supplementary Tables .....	188
3.8.5	Supplementary References.....	194

<b>4</b>	<b>General Discussion .....</b>	<b>197</b>
4.1	Summary of Performed Studies and Main Findings .....	198
4.2	A Framework for Studying the Computational Mechanism of Route and Map-based Navigation.....	200
4.2.1	Cognitive Processes Pertaining to Route-based Navigation .....	201
4.2.2	Cognitive Processes Pertaining to Map-based Navigation.....	203
4.3	Neural Correlates of Cognitive Processes Pertaining to Route and Map-based Navigation.....	205
4.3.1	Place Recognition .....	205
4.3.2	Sequence, Response, and Associative Learning.....	207
4.3.3	Locating the Goal.....	208
4.4	Decision in Space: Where Do We Go from Here? .....	210
4.4.1	Potential Research Constraints .....	210
4.4.2	Implication for the Future of Spatial Navigation Research .....	211
4.5	Conclusion.....	215
	Reference .....	219
	List of Abstracts and Publication .....	241
	Affidavit .....	243
	List of Author Contribution .....	245

# Figures

Figure 1.1. The relationship between egocentric and allocentric reference frame .....	38
Figure 1.2. An overview of a human analog of the Morris Water Maze experiment....	51
Figure 1.3. An example of a Cross Maze task .....	52
Figure 1.4. An overview of a Radial Arm and Starmaze task.....	54
Figure 1.5. The hexatown paradigm .....	55
Figure 1.6. An overview of the Dual Strategy Paradigm (DSP) .....	55
Figure 1.7. The Medial Temporal Lobe .....	58
Figure 1.8. Schematic of the Retrosplenial Complex (RSC) .....	66
Figure 1.9. Processes in the value-based decision making .....	70
Figure 1.10. Model-based and model-free choice in a simplified maze task .....	71
Figure 1.11. The agent-environment interface and elements of reinforcement learning .....	73
Figure 1.12. The SARSA with eligibility trace algorithm .....	76
Figure 1.13. Comparison of valuation based on various SARSA algorithms .....	77
Figure 1.14. Value iteration algorithm.....	79
Figure 1.15. Major subdivisions of the frontal lobe .....	82
Figure 1.16. Basal ganglia .....	86
Figure 2.1. Wayfinding task and behavioral results .....	118
Figure 2.2. Reinforcement learning models and model fits .....	119
Figure 2.3. Correlation between navigation indices and weight parameters (n=27) ...	120
Figure 2.4. Correlations of model predicted values with BOLD signals.....	121
Figure 2.5. Correlation between subjects' relative degree of model-based behavior and fMRI parameter estimates .....	122

## FIGURES

---

Figure 3.1. Wayfinding task and examples of participants' path .....	169
Figure 3.2. Reinforcement learning models and model fitting.....	171
Figure 3.3. Behavioral Results .....	173
Figure 3.4. BOLD activity pertaining to model-free value signals .....	174



# Supplementary Figures

Supplementary Figure 2.1. Example of five representative subjects' paths in encoding and search phases .....	138
Supplementary Figure 2.2. Correlation of navigation indices with parameter weight on every phase of the wayfinding task .....	139
Supplementary Figure 2.3. Distribution of fitted parameters for the behavioural and fMRI experiments.....	140
Supplementary Figure 2.4. Distance and model-based regressor in one GLM.....	141
Supplementary Figure 3.1 Path during the training phase.....	186
Supplementary Figure 3.2. Influence of gender on spatial ability .....	187



# Supplementary Tables

Supplementary Table 2.1. Navigation Indices, Fitted Parameter Weight ( $\omega$ ), and MRT Score.....	142
Supplementary Table 2.2. Distribution of subjects' individual maximum likelihoods and parameter estimates.....	143
Supplementary Table 2.3. Activation at decision points for model-free valuation across three phases of the wayfinding task.....	144
Supplementary Table 2.4. Activation at decision points for model-based valuation across three phases of the wayfinding task.....	145
Supplementary Table 3.1. Navigation Indices, MMSE Score, Fitted Parameter Weight, and MRT Score of the Older Participants.....	188
Supplementary Table 3.2. Navigation Indices, MMSE Score, Fitted Parameter Weight, and MRT Score of the Female Younger Participants.....	189
Supplementary Table 3.3. Navigation Indices, MMSE Score, Fitted Parameter Weight, and MRT Score of the Male Younger Participants.....	190
Supplementary Table 3.4. Results of Model Fitting for the Older Participants.....	191
Supplementary Table 3.5. Results of the RL Model Fitting for the Male Younger Participants.....	192
Supplementary Table 3.6. Activation at decision points for model-free valuation across three phases of the wayfinding task.....	193



# Abbreviation and Symbols

AD	Alzheimer's Disease
BA	Brodmann Area
BOLD	Blood Oxygen Level Dependent
dIPFC	dorsolateral prefrontal cortex
EPI	Echo Planar Imaging
fMRI	Functional Magnetic Resonance Imaging
FWHM	Full Width at Half Maximum
GLM	General Linear Model
ICA	Independent Component Analysis
MCI	Mild Cognitive Impairment
MMSE	Mini Mental Status Examination
MNI	Montreal Neurological Institute
mPFC	medial prefrontal cortex
MRI	Magnetic Resonance Imaging
MTL	Medial Temporal Lobe
PET	Positron Emission Tomography
PFC	Prefrontal Cortex
RAM	Radial Arm Maze
ROI	Region of Interest
RSC	Retrosplenial Cortex
SD	Standard Deviation
SNR	Signal to Noise Ratio
SPM	Statistical Parametric Mapping

VE	Virtual Environment
vmPFC	ventromedial prefrontal cortex
VR	Virtual Reality
$\alpha$	alpha, the learning rate
$\beta$	beta, temperature parameter, randomness in choice
$\lambda$	lamda, eligibility trace
$\gamma$	gamma, discount rate
$\omega$	omega, weight parameter, propensity of model-based







# General Introduction

Most of us can travel to and from work every day with little to no problems. Some of us even arrive at particular destination with little recollection of the path taken. The ability to find our way through a large-scale space is undoubtedly essential for successful function in the modern world. Consequently, spatial navigation has inspired a great deal of research, including neuroimaging studies of humans and single unit recordings of animals. Despite its importance, we often take our ability to navigate for granted. We tend to ‘ignore’ this ability until we get lost in a new city or when it is compromised by old age.

Furthermore, successful navigation recruits wide range of cognitive abilities including sensory processing, memory, and executive functions. This process is also informed by spatial knowledge derived from internal and external cues. Recent research has been particularly notable for increased understanding of not only the factors affecting human navigation but also the nature and properties of navigation strategies: route-based and map-based navigation (Wolbers and Wiener, 2014; Ekstrom et al., 2017).

On the one hand, the route-based navigation, which allows habitual traversal of fixed routes (Golledge, 1999; Latini-Corazzini et al., 2010; Chersi and Burgess, 2015), relies on memory for sequences of distances and turns. On the other hand, the map-based strategy requires the encoding of spatial relationship between goals, landmarks, or other salient points in space (Golledge, 1999; Wolbers and Büchel, 2005; Nadel, 2012; Ekstrom et al., 2014). Although computationally more demanding, the map-based navigation enables planning and selection of novel paths.

In most neuroimaging studies on spatial navigation (Aguirre and D'Esposito, 1997; Iaria et al., 2003; Bohbot et al., 2004; Wolbers et al., 2004; Wolbers and Büchel, 2005; Epstein and Higgins, 2007; Foo et al., 2007; Iaria et al., 2007; Doeller et al., 2008), the route-map dichotomy has been investigated in two different environments, two different trials, two different perspective encodings, or two different groups of participants, one for each kind of navigation strategy. These methodological choices offer the advantage of reduced risk of interference between the two strategies and thus highlight the two distinct networks involved in either route or map-based navigation. In the case of interactions between route-and map-based navigation, however, a comprehensive approach should consider that in everyday life these two strategies might work in parallel. Consequently, although there is a growing number of studies reporting neural correlates of spatial processing, the computational mechanism accounting for interaction of route and map-based navigation as well as its neural underpinnings remains a topic of debate, particularly in humans.

One promising domain to dissect the computational mechanism of spatial navigation is value-based decision making. Similar to spatial navigation, there is a prominent view in value-based decision making that the brain engages two complimentary systems for value-based decision: model-free and model-based choices (Rangel et al., 2008). The former works by reinforcing successful actions and avoiding task structure. The latter, in contrast, relies on a cognitive representation of the task. This representation is used to evaluate which set of actions lead to the best outcome. Extensive studies have shown that the computational principles underlying these two value-based-choice systems are well explained by a number of reinforcement learning algorithms. In addition, combining reinforcement learning model and neuroimaging methods, studies also report how these two complimentary systems coexist in different corticostriatal circuits (Dayan and Niv, 2008; Hare et al., 2008; Glascher et al., 2010; Daw et al., 2011; Wunderlich et al., 2011; Glascher et al., 2012; Wunderlich et al., 2012).

This doctoral thesis focuses on bridging the gap between two rather separate fields: spatial navigation and value-based decision making. Using a novel wayfinding paradigm in a Virtual Reality (VR) setting and computational model based on reinforcement learning (RL), the studies presented in this thesis, tested whether the model-free choice that rests on forming association between stimuli and responses is comparable to route-

based navigation that is based on linking cues in the environment with sequence of movement and landmarks. Likewise, model-based choice is performed by searching down a task structure that is notably similar to a spatial map. Moreover, by measuring blood oxygen level dependent (BOLD) signal while human subjects perform the wayfinding task, this thesis explore the idea that neural processing of reward based information and spatial navigation interact closely. Thus, they might share common neural networks. The reason for this hypothesis is that most navigation tasks require several decision processes to be solved.

Based on the objective of this doctoral thesis, this chapter positions the thesis in relation to existing research. The literature of a number of pertinent issues, spanning different academic fields, is reviewed. Two broad areas are discussed: (1) spatial navigation and wayfinding, and (2) value-based decision making. The first half of this chapter will focus upon the body of literature concerned with strategy adoption in spatial navigation. It will begin by discussing the concept and definition of spatial navigation and wayfinding. Through survey of definitions, this chapter will gradually distil a working definition of two strategies in navigation. Next, studies highlighting different methodologies in assessing navigation strategies are reviewed. Of particular interests are those attempting to compare real and virtual environment. This chapter then moves on to highlight neural basis of navigation strategies. The second half of this chapter will focus on value based decision making. Of particular interests are studies dissecting different strategies in value-based decision making as well as neural correlates of reinforcement learning as a computational framework to represent choice behavior. This chapter then concludes with the aim of the thesis.

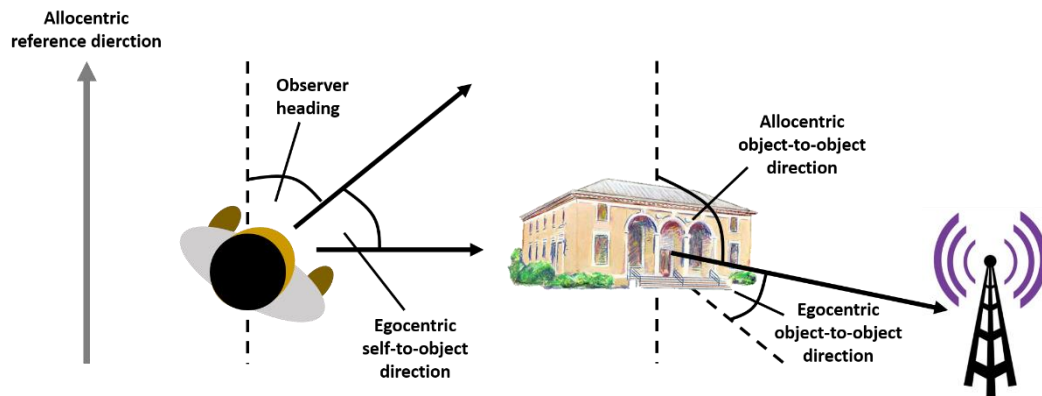
## 1.1 The Concept of Wayfinding and Spatial Navigation

Before a deeper explanation into the topic of this thesis, it is helpful to define and clarify some of the basic assumptions, ideas, and terms that will be presented throughout this thesis. There are at least two ways to describe the processes behind human movement in a large-scale environment namely wayfinding and spatial navigation. Although in most literature the behavioral, cognitive, and neural process involved during navigation and wayfinding are typically used interchangeably, there are some crucial discrepancy between these processes. The first term, wayfinding, refers to the behavior of finding one's way from an origin to one or more destinations (Heft, 2012). It involves selecting path segments from an existing spatial representation and evaluating these segments against previous spatial decisions as one travels along a specific path. The cognitive components of wayfinding center on the ability to know origins and seek destinations that may have never been visited, to determine turn angles in appropriate sequence, to maintain orientation, as well as to estimate location based on landmarks (Montello, 1991).

Furthermore, wayfinding is a purposive, directed and motivated activity that relies on existing spatial knowledge. This spatial knowledge is gathered from prior experience, maps, or even verbal descriptions. An example of individuals with a phenomenal wayfinding ability would be experienced taxi drivers. Studies have demonstrated that experienced taxi drivers can generate novel routes and remember street names better than pedestrians and bus drivers (Maguire et al., 2000). These findings are even supported by corresponding hippocampal neurogenesis that is suggested to reflect the development, storage, and use of a complex spatial representation (Spiers and Maguire, 2006).

The second term, spatial navigation, is used as a proxy for the processing of a variety of different forms of information regarding position and rate of travel between identifiable origins and destinations summarized as a course to be followed. The etymological roots of navigation come from the Latin word *navis*, which refers to the art of travelling, often by sea. Colloquially it means to deliberately walk or make one's way through some space. In contrast to wayfinding, the focus of spatial navigation is on the action rather than the cognitive processes behind it.

The information used in navigation is most often visual but also vestibular, proprioceptive, somatosensory, and auditory (Wolbers, 2010). Although humans have a bias toward using visual information, vestibular input and the efference copy of self-motion are extremely important for real-life navigation. Together, they all contribute (either in a combined fashion or independently) to extracting information about the environment (e.g., its shape and scale), the location of items, and our own location within it (Holden and Newcombe, 2012). While navigating, we become familiar with the environment and acquire knowledge about it, thereby extracting information from it and storing this information in our memory. This way, we can recall it later for a variety of purposes (Golledge, 1999). Based on these definitions, this is the distinction between wayfinding and spatial navigation to be used in this thesis.



**Figure 1.1. The relationship between egocentric and allocentric reference frame**

In an egocentric coordinate system locations are represented relative to the body-orientation of a navigator as indicated by the arrow. An allocentric reference frame is located and oriented on an object or a location other than the navigator. Stationary objects do not change their coordinates or bearings when the navigator (i.e., the representing system) is moving. Distance, direction and bearing of an object relative to the origin of the coordinate system are directly represented. Figure adapted from Wolbers and Wiener, 2014.

### 1.1.1 Reference Frames in Wayfinding and Spatial Navigation

In order to interact with space, during wayfinding and spatial navigation, humans need to integrate spatial information to new objects encountered and new motion information. This process is supposed to rely on reference frame (Gomez et al., 2014). This is because a reference frame determines what kind of spatial environment is represented, and how locations are specified within that environment (McCloskey, 2001). Exceptions notwithstanding, there is a general agreement that there are two kinds of spatial reference frames (as depicted in Figure 1.1): egocentric and allocentric (Wolbers and Wiener, 2014).

This doctoral thesis refers to egocentric or ego-relative spatial reference frames whenever locations are represented with respect to the particular perspective (be it of the present, remembered, or imagined) of an observer, as opposed to an external landmark (Filimon, 2015). The origin of the egocentric reference frame is centered on the observer. Thus, its orientation is defined by the observer's heading.

Furthermore, there are some assumptions in defining egocentric reference frame including: (1) all body parts of the observer are oriented in the same direction, (2) the observer's heading coincides with the orientation of all egocentric reference frames, and (3) polar coordinate system is used (Wolbers and Wiener, 2014). Based on these assumptions and as shown in Figure, egocentric distance can be defined as the length of

vector connecting the observer and the object. The angle between the observers' heading and that vector specifies its egocentric direction (Taylor and Brunye, 2012).

Allocentric reference frame, also known as object-oriented or view-independent, is generally used to include any form of spatial reference where locations are encoded relative to points external to the observer. As shown in Figure 1.1, an object's allocentric distance corresponds to the length of a vector connecting the origin of the coordinate system and the object. This distance is rarely defined with respect to an origin but rather to local representation of relative, object to object spatial relationships. Similarly, allocentric direction usually refer to the direction between two external objects (Taylor and Brunye, 2012; Wolbers and Wiener, 2014; Ekstrom et al., 2017).

Based on these definitions, it is clear that the differences between egocentric and allocentric reference frames become crucial when the observer moves about in the environment (i.e. during wayfinding and navigation). For example, when an observer simply turns around, only object's egocentric direction changes while its egocentric distance remains the same. In contrast, neither allocentric distance nor direction is affected by observer rotation. These differences are more pronounced for translational movements in which both allocentric distance and direction remain unaffected.

### **1.1.2 Organization and Properties of Spatial Knowledge**

Spatial knowledge, as the term used in this thesis, is knowledge of the locations of objects, places, and environmental features. This knowledge, which underpins navigation through an environment, can be classified into three distinct types: object-place, route, and graph/survey knowledge (Siegel and White, 1975). The object-place knowledge is the knowledge of the identities and appearance of entities. It is usually the first to be acquired and become the building block of other types of spatial knowledge. Some of these entities are considered objects e.g. coffee table, whereas others correspond to significant location of greater extent and less well-defined boundaries e.g. small city park. Landmarks are entities of special significant to wayfinding and navigation. These landmarks might be used as: (1) indicator of locations of other objects and places (e.g. the research center is at the top floor of the university hospital), (2) goal

of navigation (e.g. I am going to the post office), and (3) marking of decision points (e.g. turn left at the university main building).

The second one, route knowledge, is typically associated with an egocentric reference frame. It consists of familiarity with sequences of landmarks and associated decisions and actions (Siegel and White, 1975). Actions specify the steps to get to the next landmark on the route. In route knowledge, the function of landmarks is comparable to associative cues. Route knowledge is assumed to be nonmetric early in acquisition, consisting only of the order of landmarks and associated actions. Through experience route knowledge can acquire metric properties and specify distances, temporal durations, and turning angles (Thorndyke and Hayes-Roth, 1982b).

The most sophisticated form of spatial knowledge is survey knowledge (also frequently compared to a 'map' or a 'graph', the 'top-down' knowledge), which is associated with an allocentric reference frame. It is the knowledge of the overall spatial layout of an environment, and includes Euclidian distances as well as inter-point directions defined in a common reference system (Hegarty and Waller, 2005). A key characteristics of survey knowledge is that the spatial relations between locations can be inferred even if travel has never occurred between locations. As a consequence, behaviors that pinpoint the presence of survey knowledge may consists of the abilities to create efficient path e.g. taking shortcuts or to pinpoint to unseen location (Golledge, 1999; Nadel, 2012).

Furthermore, spatial knowledge has several key properties including fragmented, distorted, hierarchical, and orientation dependent. First, spatial knowledge is typically fragmented. This means it consists of a patchwork of detailed knowledge of some areas and only spare knowledge of other areas (Appleyard, 1970). A second key property of spatial knowledge is that memories of spatial relations, such as distances, angles, and orientation, often differ from the physical values in systematic and predictable ways. Example of these distortions include greater estimates of Euclidian distances when locations are separated by boundaries (Wiener et al., 2009), asymmetric distance estimation under different task instructions (Takahashi et al., 2013), and angles of intersection between roads are remembered to being closer to  $90^0$  than they are in reality (Allen, 2006). Moreover, there is strong evidence that memories of locations of objects in the environment are organized categorically and hierarchically. Consequently, a region



of space may be represented as a whole and as a part. The former means that a region of space contains other regions and locations. On the contrary, the latter means that a region of space is contained in a larger regions. Lastly, humans recall and recognize spatial relations between objects more efficiently from some perspectives than from others. Thus, spatial knowledge is known to be orientation dependent.

## 1.2 Strategy Adoption in Spatial Navigation and Wayfinding

As written in the beginning of this chapter, successful navigation may rely on two types of strategy adoption, generally termed as route-based and map-based navigation. Route-based navigation involves recognizing an origin and a destination, and identifying route segments, turn angles, and the sequence of segments and angles that make up the desired path. In contrast, map-based navigation requires encoding of the spatial relationship between landmarks and/or locations. It relies on a cognitive map that supports flexible navigation within an environment. Building from this idea, much of this chapter is an elaboration of the characteristics of these two strategies and how normal aging affect strategy adoption in navigation.

### 1.2.1 General Concept of Route-based and Map-based Navigation

The major concern of route-based navigation is using the pattern of the path and the sequence of behaviors needed to traverse it, rather than learning the environment through which the route passes. Thus, it relies mainly on route knowledge and the egocentric spatial reference frame. When route-based navigation takes place, on-route information is dominant and take precedence over all off-route information. Environmental features such as landmarks are learned to support the acquisition of spatial knowledge at each segment of the path taken. This process is employed in a manner of stimulus-response pairings that involves relating a directional response with an encoded landmark at certain segments of the path taken. The behavioral response associated with successful route-based navigation are encoded relative to one's body and thus, route-based navigation relies mostly on egocentric reference frame. Consequently, it supports accurate navigation when the position and orientation in the environment are known.

Furthermore, environmental features that are spatially correspond with a goal location can be encoded as beacon. Subsequent recognition of encoded landmarks triggers a universal behavioral response that results in movement relative to the position of the landmark such as 'turn right toward the train station'. This recognition of the encoded landmark during subsequent navigation facilitates the recall of the corresponding route knowledge, which is then used to inform spatial behavior.

Map-based navigation relies on cognitive map, a concept first introduced by Tolman in 1948. Although the notion of a cognitive map has been used in many different ways since then, in this thesis, a cognitive map is defined as a mental representation of the environment that captures the spatial relations among geometrical structures in a defined region (Tolman, 1948; Golledge, 1999). Especially in human and other primates, the term cognitive map implies deliberate and motivated encoding of environmental information. This information can be used to: (1) determine where one is at any given moment, (2) locate where specific encoded objects are in surrounding space, (3) specify how to get from one place to another, and (4) communicate spatial knowledge to others. Note that it is not necessary to assume that a cognitive map looks like a map in the head. Nevertheless, it is commonly agreed that a cognitive map consists of points (such as landmarks and reference node), lines (such as routes and tracks), areas (for example, regions and neighborhoods), and surfaces (three dimensional characteristics of places) (Golledge, 1999; Nadel, 2012).

There are three key features of a cognitive map. First, it represents information in a viewpoint-independent fashion. Another key feature is that it represents configurations, rather than simple associations, between features of the environment. This configuration of features can be attained in a variety of ways including defining boundaries around a certain region, integrating separately learned route information into a network, and overviewing from a survey point. Finally, a cognitive map depend on allocentric reference frame (Thorndyke and Hayes-Roth, 1982a; Maguire et al., 1999; Foo et al., 2005; Smith and Mizumori, 2006; Iaria et al., 2007). There has been considerable debate about the relative importance and roles of egocentric versus allocentric reference frames in the formation of a cognitive map, and it is beyond the scope of this chapter and this thesis to join the debate.

Given that a cognitive map represents environmental realities, it is acquired through interactions with the external world. These interactions can take, at least, three forms: (1) active search and exploration according to specific rules or heuristics, (2) a priori familiarization with secondary information sources about the environment, and (3) experience of the environment using controlled navigational practices (Golledge, 1999). Through these forms of interactions, various details of the environment are captured. The more interactions take place, the more details are acquired. Over some period of

time, given that one gains more exposure to the same environment, cognitive map of a given environment becomes progressively more detailed.

### **1.2.2 Age-related Changes in Navigation Strategies**

Older age is associated with functional decline in selective aspects of cognitive performance including executive function, attention, verbal and visual explicit memory, as well as working memory (Park et al., 2001). Given that spatial navigation involves a multitude of cognitive functions and processes, it is not surprising that there is an age-related preferential shift for one navigation strategy over the other. This preferential shift has been reported by a steadily accumulating literature investigating age related navigation differences in humans. This subsection reviews the extant literature in three categories: (1) studies that pointed out age-related preference for route-based to map-based navigation, (2) studies that underlined age-related impairments in utilizing allocentric reference frame, and (3) studies that highlight the age-related difficulties in switching between reference frames.

At least two experimental studies reported an age-related preferential use for route-based rather than map-based navigation (Rodgers et al., 2012; Goeke et al., 2015). By comparing the performance of older to younger adults in a virtual version of the Morris Water Maze (hMWM, details in section 1.3), Rodgers et al. (2012) reported that older adults achieved worst results. They travelled longer distance before finding the target location. Moreover, comparing the two groups on their preferential strategy through a virtual Y-maze task (details in section 1.3), the study found that older adults used route-based navigation more often. The younger adults were almost equally distributed with 46% preferred to use route-based navigation. In contrast, the older adults were more likely to adopt route-based navigation (82%). Similar age-related preference was reported by Goeke et al. (2015) using the tunnel paradigm. These authors concluded that age-related alterations in the neural system supporting map-based navigation, especially in the medial temporal lobe (MTL), may drive the elderly to more frequent use of route-based navigation.

Gazova et al. (2013) assessed the progressive deterioration of spatial reference frames among different age groups. The authors reported a quadratic effect of age and the ability of using allocentric reference frame (Gazova et al., 2013). Specifically, the allocentric decline was mainly observed at subjects older than 70 years of age. In line with this result, Wiener et al. (2012) showed that older adults have difficulties in route retracing. This indicates that older adults cannot indicate direction from the opposite site of the original learned route. In contrast, older adults performed reasonably well during the route learning task where they had to repeat the same learned route (Wiener et al., 2012). Along with findings from Moffat et al. (2006), these studies provide evidence toward specific allocentric, but not egocentric, impairments that may reflect age-related hippocampal decline.

Furthermore, successful navigation depends largely on the ability to integrate information from different spatial reference frames. This ability is also affected by aging because older adults have difficulties in switching between reference frames. In both of their studies, Harries et al. (2012; 2014) reported that aging has a major effect on allocentric to egocentric translation. They did not, however, find any age-related impairments on the use of allocentric reference frame nor on the egocentric to allocentric translation (Harris et al., 2012; Harris and Wolbers, 2012). Contrary to these findings, two studies identified age-related declines in the opposite direction (Carelli et al., 2011; Morganti and Riva, 2014). Both studies used the Wisc-R papers and pencil (P&P) maze task and an equivalent virtual version of the task. To solve the virtual version of the task, survey knowledge from the P&P version needs to be translated into an egocentric frame. Older adults performed worse, especially in the virtual version of the task. This effect was strongly evident as age increased and positively correlated with the Mini Mental State Examination's score (MMSE).

### **1.3 Assessing Wayfinding Strategies Using Virtual Environment**

One fundamental question that is generally considered when assessing strategies in spatial navigation and wayfinding is how the environment will be presented during the assessment. Traditionally, most navigation studies are carried out in the so called real world. This setting offers, in addition to convenience, the possibility to produce patterns of behavior similar to daily life as a significant advantage (Sitzmann et al., 2018). Nevertheless, real world setting poses a number of challenges. Contrary to studies in a laboratory setting, influencing and constraining a real world environment can be challenging. Enforcing identical conditions for all participants or optimize task design might be influenced by potential disturbing factors such as weather conditions, traffic, noise (Gillner and Mallot, 1998; van der Ham et al., 2015). Furthermore, participants with prior knowledge of the test environment add experimental noise to the measurement. These challenges along with the rapid improvement of technology, particularly with regards to virtual environments, promote the use of virtual reality technology as a prominent tool in navigation research. Since the early 1990's, when computers became powerful enough to simulate large-scale environment, the growth of papers in this topic has been exponential. This is largely due to the fact that with these virtual environments, wayfinding strategies can be studied in any possible type or set up of environment. In addition, behavioral responses can be recorded in great detail. In this section, studies comparing navigation in real and virtual environments as well as those proposing various set up to compare route and map-based navigation are reviewed.

#### **1.3.1 Navigation in Virtual vs. Real Environment**

As mentioned in the beginning of this section, real world environments might arguably produce patterns of behavior that better reflect how humans navigate on a daily basis. In some cases, however, virtual environments can better facilitate the isolation of different processes that construct navigation. In this subsection we provide a brief overview of the benefits and challenges associated with VR in navigation research. We discuss its utility in three ways: (1) discerning advantages of VR over conventional stimulus presentation, (2) understanding the technological limitations of the methodology, and (3) investigating whether skills or knowledge can be carried over from the real world to the virtual world and vice versa.

To begin with, the term VR is often used interchangeably to refer to one of three systems: (1) a virtual environment shown on a flat screen, (2) a room-based system with sensors attached to one's body, or (3) a head-mounted display (Wilson and Soranzo, 2015). Though all three systems are quite different, a common feature of all three is the presence of depth. This feature creates an illusion that the viewer is seeing objects in a virtual space (Diemer et al., 2015). There are a number of advantages offered by this feature including greater control over stimulus presentation, variety in response options, and possibility of increasing ecological validity.

There are several advantages in presenting experimental stimuli using VR technology, especially in spatial navigation and wayfinding research. First, VR can present a range of complex stimulus that would not be easily controllable in the real world and enabling the examination of both cognitive process such as attention and functional behaviors such as planning a series of actions along a path in a virtual maze (van der Ham et al., 2015; Grubert et al., 2017). Consequently, in VR participants are enabled to respond to pertinent stimuli while being immersed in a larger environment which can itself be controlled (Waller et al., 2003). This differs from traditional experimental contexts where the pertinent stimuli may be controlled but the surrounding environment often cannot be. Second, virtual environments can be modeled and controlled to the experimental requirements, without having to build something similar in the real world (Dombeck and Reiser, 2012). Consequently, situations that would be impossible in the real world, such as teleportation, can be realized in a VR setting. Third, VR enables the development of new protocols to measure participant responding that goes beyond point and click exercise. This is especially important in navigation research where researchers need to strike a balance between controlled environment and ecological validity. Lastly, VR allows studies to be conducted in a lab setting. This means experimental conditions can be further constrained and remained comparable for all participants. This high level of control improves the validity and might improve the reproducibility of navigation studies (Schultheis et al., 2002).

Despite the advantages of VR in navigation research, it is important to consider to what extent the technology itself affect the outcome of the experiment. These technological issues might arise from interface design, control of motion through the environment, and field of view of VR displays. Ruddle et al. (1997) used direction and distance tests

to determine whether there is a difference in patterns of movement between subjects navigating immersively and non-immersively. In this study, subjects used either a head-mounted (immersive) or a desktop display (non-immersive). On the whole, they found that the proprioceptive feedback for the group using headset did not drastically improve their orientation judging ability. As a results, there were similar patterns of movement between immersive and non-immersive group (Ruddle et al., 1997).

Following the work by Ruddle et al. (1997), Wells et al. (1996) tried to address the effects of navigational control devices. In particular, whether navigational control devices that used whole body movements are superior to control devices that are independent of body positions (such as joysticks). Using route learning and direction orientation task, they found that in simple environments, there was a negligible difference between devices. However, the body controlled device were superior as environments became more complex. Some other studies have also dealt with this topic and found that, in a complex environment, the performance of body controlled device is merely superior (Peterson et al., 1998).

As for the effects of field of view (FOV), there has been some conflicting results (Bakker et al., 2001). Using headset of differing FOV and a number of homing tests, Peruch et al (2001) concluded that the amount of simultaneous environmental information is not a determiner of one's comprehension of the environments. On the contrary, earlier work by Alfano and Michel (1990) found that restriction in FOV affect the performance of certain tasks related to navigation and wayfinding. These tasks include following a winding path, hand-eye coordination, and forming a mental representation of the environment (Alfano and Michel, 1990).

Lastly, the question remains whether the results of VR studies are valid in comparison with traditional navigation research methods. Therefore, the final part of this subsection is dedicated to review what is known about the degree to which navigation behavior in VR and real environment is comparable. An early work by Witmer et al. (1996) examined how route knowledge gained in a simulation of a complex building can aid navigation in the real building. Comparing performance of three groups (route in real building, virtual simulation of the building, and color photographs), they measured subjects' spatial knowledge acquisition and wayfinding ability. Their results suggested that virtual



environments that adequately represent real world complexity can be effective training media for learning complex routes in buildings (Witmer et al., 1996).

In the same year, Thaulka and Wilson (1996) tested whether knowledge gained in the virtual world was flexible and independent of the orientation of the observer. There were two groups of participants: one group gained spatial knowledge through virtual environments, the other through examination of a map of the same environment. They found that while the group that had studied the map had orientation-specific knowledge of the world, the group that had navigated through the virtual environment had an orientation-free knowledge of the test environment. In other words, they suggested that real-world and simulated navigation both result in similar orientation-free cognitive representation of the environment. In addition, in accordance to previous work, they concluded that there is a great deal of equivalence of learning in simulated and real space (Richardson et al., 1999).

Following these works, there have been several studies with a positive outlook towards comparability of navigation in real vs. virtual environments. These studies often perform a navigation study in both virtual and real environments and compare the results between groups. They find that participants are able to effectively navigate through virtual environments and are able to learn spatial relations in the virtual environment. Combining these findings with the promising features VR offers, for instance high controllability of stimuli and precise measurements, these studies concluded that VR can be successfully employed in assessing various characteristics of navigation and wayfinding behavior in humans (Gillner and Mallot, 1998; Schmelter et al., 2009).

Despite these promising findings, several studies indicate that even though participants are indeed able to acquire some navigational knowledge from a virtual environment, navigation performance and spatial learning simply do not reach the same level as real world conditions. Their opinion about the application of VR are often more nuanced. Correlations between how well participants were able to learn virtual and real environments are found, as well as similar biases between the two groups. This seems to suggest that similar cognitive mechanisms are being used in virtual learning. Despite these promising findings, these studies also indicate that virtual environment learners

simply showed the poorer learning, than real world learners (Richardson et al., 1999; Waller et al., 2004; Schmelter et al., 2009).

All in all, the majority of studies comparing navigation in real and virtual environments appear to be suggesting that we use real and virtual space analogously. In other words, knowledge gained in either one may be applied to the other. This conclusion clearly has many applications, especially with regards to assessment of wayfinding strategies. In the next segment of this section, different set ups of virtual environments that facilitate the investigation of wayfinding strategies will be presented.

### **1.3.2 Assessment Paradigms in Virtual Environment**

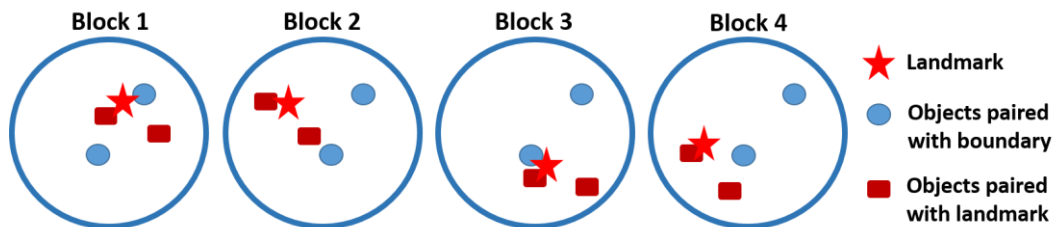
Because there are many advantages of using VR in spatial navigation research, a number of researchers started developing different assessment paradigm with the help of VR. Some of these researchers focused on the effect of technology and interface. Others performed wayfinding experiment by simply substituting virtual world for the real world. The one presented here are those that specifically target the strategy adoption in navigation.

#### ***The Morris Water Maze Task***

The most widely used paradigm for assessing wayfinding strategies is the Morris Water Maze Task (MWMT). It was originally developed by Richard Morris to examine the spatial memory of rodents. It consists of a circular pool containing opaque water and a hidden submerged target platform without the presence of proximal cues. Rodents are placed in the pool and rely on distal cues to find the platform. As the rodents became more familiar with the task, they can find the platform quicker. Several characteristics contributed to prevalent use of MWMT. These include lack of required pre-training phase, reliability across a wide range of tank configuration and testing procedures, as well as extensive evidence of validity as a measure of hippocampal-dependent map navigation (deBruin et al., 2001).

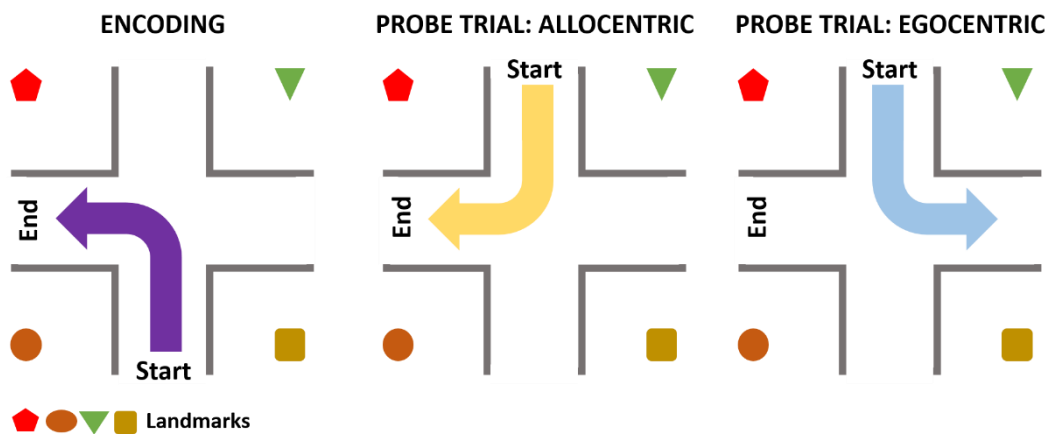
The task has also been adapted in both laboratory and VE settings to test human participants (Astur et al., 2002; Moffat and Resnick, 2002; Driscoll et al., 2005;

Newhouse et al., 2007; Hamilton et al., 2009; Herting and Nagel, 2012). One example is the work by Doller et al. (2008). Using VR technology, they adapted MWMT into an object-location memory task (Figure 1.2) comprising a landmark, a circular boundary (analogous to the water tank), and distant cues. In the task, some objects maintained a fixed location relative to the environmental boundary. Other objects maintained a fixed location relative to a single landmark. Participants learned the locations of the objects by collecting and replacing them over multiple trials while being scanned by fMRI. The distinct incidental hippocampal activation suggest the use of cognitive map (Doeller et al., 2008).



**Figure 1.2. An overview of a human analog of the Morris Water Maze experiment**

Participants learned four object's locations over four blocks, the landmark (red star) and boundary (large blue circle) moving relative to each other at the start of each block. Two objects were paired with the landmark (blue circles), and two objects were paired with the boundary (red squares).



**Figure 1.3.** An example of a Cross Maze task

The left most panel is an overview of the training or the encoding phase. Participants learn to navigate from a fixed starting position to a goal arm. The middle and right panels are overview of the probe trial. The middle panel illustrates how subjects' response when subjects use allocentric reference frame while the right panel illustrates subjects' response when they use egocentric reference frame.

### ***The Cross Maze Tasks***

The Cross Maze (Tolman, 1948; Schroeder et al., 2002; Wingard and Packard, 2008) was originally developed to investigate navigation behavior in rodents, and have since been adapted to study human spatial cognition. As shown in Figure 1.3, this paradigm features four arms radiating from a central junction. Experiment employing cross maze usually consists of a sequence of training and probe trials. During training, participants navigate from a fixed starting position, usually at the end of a radial arm to a goal arm located either to the left or right of the maze junction. Use of map-based navigation in this maze requires the knowledge of the spatial relationship between the goal arm and distal environmental cues. On the contrary, route-based navigation only involve a simple *turn left/right* response. Strategy preference is assessed by probe trials. During these trials, the starting position is different to that in training, usually shifted to a different radial arm.

One example of the use of Cross Maze Task in VE setting is the work by Wiener et al. (2013). The authors used Cross Maze to construct a route-learning paradigm that investigates the effects of cognitive aging on the selection and adoption of navigation strategies in humans. Since participants were required to rejoin a previously learned route encountered from an unfamiliar direction, navigation to the position of the goal suggests presence of cognitive map (Figure 1.3). Likewise executing the same turning response as the one learned during the encoding phase is indicative of route-based navigation (Wiener et al., 2013).

### ***The Radial Arm Maze (RAM) and Starmaze Paradigm***

Initially, the Radial Arm Maze (RAM) was employed to study spatial working memory in rodents (Chen et al., 1994). However, it has been recently modified to discriminate between map-based and route-based navigation. There are several variants of this task with the most common uses eight identical arms radiating outward from a circular center platform (Figure 1.4. A). Distal cues are placed surrounding the environment. Following an encoding phase, participants' strategy preference is determined by the rate of navigational errors during subsequent probe trials, in which all available distal cues are obscured from view. As discussed in Subsection 1.2.1, map-based strategy requires the knowledge of spatial relationship among landmarks and other environmental features. Therefore, the absence of distal cues is partially detrimental to the use of map-based strategy (Iaria et al., 2003).

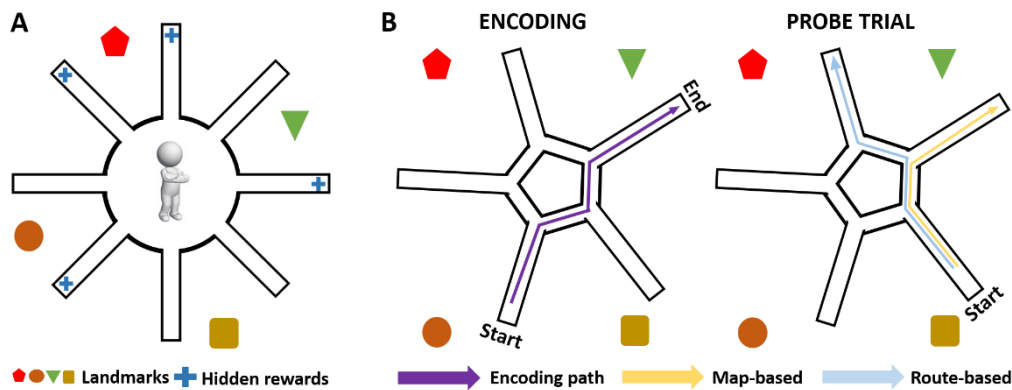
The Starmaze paradigm (Figure 1.4. B) is constructed of 10 identical alleys. Five of these alleys form a central pentagon. The remaining alleys radiate from the vertices of the pentagon. Similar to the RAM task, Starmaze task usually consists of encoding and probe trials. The objective of training trials is learning to navigate from a fixed starting position in one of the radial alleys to a goal located in a separate radial alley. To successfully navigate, participants can employ one of two strategies. On the one hand, map-based navigation by relying on allocentric reference frame to learn the location of the goal arm relative to surrounding distal cues. On the other hand, a sequential route-based navigation by relying on egocentric reference frame to learn the correct series of movement required to navigate to the goal arm.

### ***The Hexatown Paradigm***

Wolbers et al. (2004) adapted the *hexatown* environment (Figure 1.5) developed by Gillner and Mallot to investigate the emergence of route knowledge and dissociate the contribution of retrosplenial and hippocampal region to the formation of a cognitive map in humans. This particular design has the advantage of allowing a route to be planned that covers the entire town by passing through every road section only once in each direction. As the roads were flanked on both sides by brick stone walls, participants' view were restricted to the immediate straight direction. Granted that, remote road

sections or intersections were invisible. At four intersections, twelve distinct buildings were placed to serve as landmarks.

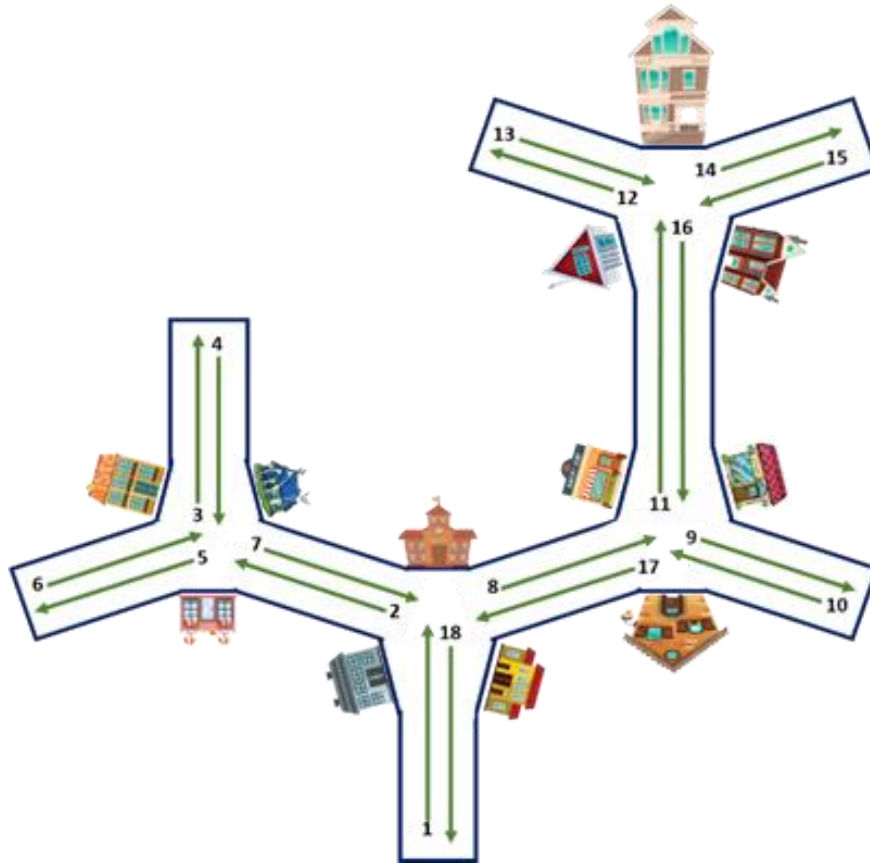
During the encoding trials, participants were moved through the environment and instructed to memorize spatial relation between landmarks. During the probe trials, participants were asked to assess spatial position of one landmark (referred as target building) relative to another landmark. Results showed evidence of emergence of route knowledge in some participants (Wolbers et al., 2004; Wolbers and Büchel, 2005).



**Figure 1.4. An overview of a Radial Arm and Star maze task**

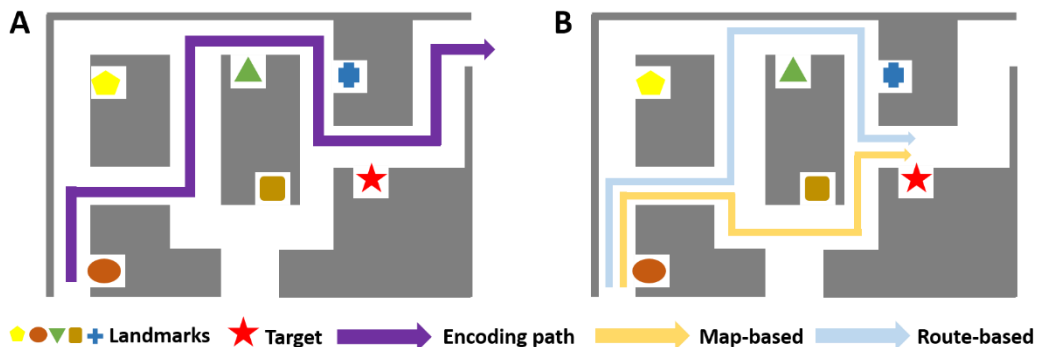
A. During training, participants learn to navigate to radial arms containing hidden rewards. Participants' strategy preference is then determined by navigation behavior during probe trials in which all landmarks are occluded from the view.

B. The left panel is an overview of the subjects' training during the encoding phase of the experiment. Subjects learn to navigate from a fixed starting position to a goal arm. The right panel is an overview of the probe trial. Subjects' responses can indicate whether they employ either map-based or route-based strategy.



**Figure 1.5. The hexatown paradigm**

The figure represents aerial view of the hexatown configuration adapted from Gillner and Mallot (1998). In the work by Wolbers and Büchel (2005), brick stone walls on both sides of the road restricted the view to the immediate straight ahead; remote road sections or intersections were invisible. Twelve distinct buildings serving as landmarks were placed at four intersections (three landmarks per intersection). These buildings were hidden behind walls unless subjects were standing directly in front of them. Stimuli for the retrieval task consisted of snapshots of all buildings from the same viewpoints as those encountered during navigation.



**Figure 1.6. An overview of the Dual Strategy Paradigm (DSP)**

*A.* A section of the training procedure. Subjects are passively transported along a route through the environment.

*B.* An overview of the probe trial. Subjects must navigate from a known location to a target landmark (red star). The use of a novel short-cut is indicative of map-based navigation use, while following the original route suggests route-based navigation.

***The Dual Strategy Paradigm***

Unlike the other four paradigms discussed in this subsection, Dual Strategy Paradigm (DSP) allow both route-based and map-based strategy adoption to support accurate – albeit different – navigation (Marchette et al., 2011). In the DSP, participants were first transported passively along a route in a grid-like virtual maze. They were asked to learn the location of twelve unique objects. During a subsequent probe trials, participants navigated to target objects from a number of locations along the original route. When participants simply navigated along the original route, this behavior suggests that they employed route-based navigation. On the contrary, when participants employed novel short-cuts, the authors concluded that the participants have developed a spatial representation of the virtual maze (Figure 1.6).



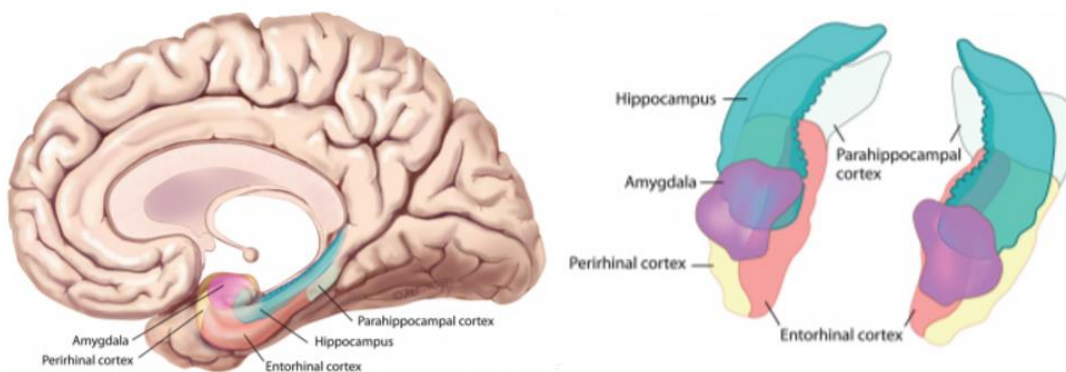
## 1.4 Neural Basis of Wayfinding Strategies

Research has shown that route-based and map-based navigation rely on different neural substrates. When it comes to addressing the fine grained questions about information encoding and processing in certain brain region, researchers typically turn to single unit recording in animals (such as rodent and non-human primate). As the name indicate, the data from single-unit-recording experiments reflect the activity of a single functional unit within a brain region that may or may not contain multiple neurons. Single unit recording uses changes in firing rate of action potential in a given neuron as an index of whether a stimulus changes the information processing with which that neuron is associated. This method involves insertion of very fine electrodes, made of metal that is highly sensitive to high frequency electrical signal, into the neural tissues adjacent to neurons of interest. Due to its extraordinary temporal resolution, single unit recording offers direct information about the rate and timing of action potentials within a brain region. Data from single unit studies provide the ground work for many subsequent studies in spatial navigation.

In humans, however, the neural correlates of spatial navigation and wayfinding are primarily investigated using non-invasive neuroimaging methods. These methods include functional magnetic resonance imaging (fMRI), positron emission tomography (PET), magnetoencephalography (MEG), and electroencephalography (EEG). Each method has its own superiorities and pitfalls. Functional magnetic resonance imaging (fMRI), which is by far the most commonly used method, measures brain activity via changes in blood flow (Huettel et al., 2004). These changes in blood flow take some time. Consequently, while fMRI offers relatively high degree of spatial precision, it suffers when it comes to temporal resolution (Logothetis, 2008). Position emission tomography (PET) uses glucose tracers to detect areas of the brain that are using the most energy. It is assumed that those areas are the most active, and thus are correlated with cognitive processes in question. Similar to fMRI, PET offers high spatial resolution but less temporal resolution. Due to the injection of tracers, it is often regarded as an invasive method. Consequently, it is not as widely used as the fMRI. In contrast to fMRI and PET, MEG and EEG offer better temporal resolution. Both MEG and EEG gauge the electrical activity (magnetic and electric field respectively) produced by the brain by

measuring neural oscillations. One limitation of both methods has to do with spatial resolution.

Those various methods have uncovered many of the primary regions involved in either route-based or map-based navigation. Briefly, the striatal circuits, which are involved in learning and enacting stimulus-response associations (Stalnaker et al., 2010), have been shown to play crucial role in route-based navigation (De Leonibus et al., 2005). In contrast, different patterns of hippocampal place cell activity across discrete areas of an environment support the notion that map-based navigation rely on the hippocampus and surrounding structure (Wilson and McNaughton, 1993). This section focuses on three functional regions that are essentials for route-based and/or map-based navigation namely medial temporal lobe, the striatal complex, and the retrosplenial complex. Whenever relevant, this section also neural representation of strategy adoption in wayfinding changes due to normal aging.



**Figure 1.7. The Medial Temporal Lobe**

The medial temporal lobe consists of the hippocampal formation (blue-green) superiorly and the parahippocampal gyrus inferiorly. The entorhinal (brown) and perirhinal (yellow) cortices form the medial and lateral components, respectively, of the anterior portion of the parahippocampal gyrus, while the parahippocampal cortex (off-white) forms the posterior portion. Adapted with permission from Raslau et al. (2014).

### 1.4.1 Medial Temporal Lobe (MTL)

Medial temporal lobe (MTL, Figure 1.7) includes a system of anatomically related structures including the hippocampal regions (CA fields, dentate gyrus and subicular complex) and the adjacent perirhinal, entorhinal, and parahippocampal cortices (Squire et al., 2004). Since the case of patient H.M., who underwent a bilateral medial temporal lobectomy for intractable epilepsy, the MTL has been associated with episodic memory. Later on, in concordance with the discovery of place cells and the idea of cognitive map, this region is thought to be involved in spatial cognition and memory in both animals (O'Keefe and Nadel, 1978; Morris et al., 1982) and humans (Maguire et al., 1998; Spiers et al., 2001).

In humans, several methods have been employed to pinpoint the role of MTL in spatial navigation and wayfinding. These methods include experimental lesion, anatomical MRI, functional MRI (fMRI), and invasive recordings. Similar to the works in rodents, findings from lesion studies suggest that damages to the human hippocampus impairs several forms of spatial cognition. Among these, the one that are crucial to strategy adoption include damage to memory for spatial relations of multiple objects in an arena (Bohbot et al., 1998), and for the spatial relation of objects within a scene (Hartley et al., 2003). In contrast to findings from hippocampal lesion, several studies showed that patients with parahippocampal lesion had profound deficits in the processing of visual-spatial information (Burgess et al., 2002; Bohbot and Corkin, 2007). These deficits, which include difficulties in learning new spatial routes and retrieving multiple new spatial routes, were not found in patients with a more exclusive hippocampal damage.

Moreover, a voxel based morphometry study showed structural differences within the hippocampal region between licensed London taxi drivers and controls (Maguire et al., 2000). These London taxi drivers must pass a rigorous test of their knowledge of London roads. In other words, they were expected to have a better spatial representation of the real world environment. On the one hand, the study found that the taxi drivers had larger posterior hippocampus. This portion of the hippocampus is crucial for both encoding and retrieval of spatial memory (Duarte et al., 2014). On the other hand, the anterior portion of the hippocampus was found to be smaller in the taxi drivers as compared to

control. This region has been suggested to be related to external attention and the dorsal attention network (Robinson et al., 2015).

In a study where strategy adoption was specifically tested, Bohbot et al. (2007) found that in a virtual RAM task, map-based navigation correlated with the gray matter density in the hippocampus. Similarly, using a virtual city paradigm, Iaria et al. (2008) found that structural integrity of the right hippocampus was correlated with how effective participants formed and used cognitive maps (Iaria et al., 2008). In line with these findings, Schinazi et al. (2013) found that as participants became increasingly familiar with a novel college campus, their performance in a pointing task was positively correlated with right hippocampal volume. The pointing task was specifically designed to rely on survey knowledge (Schinazi et al., 2013).

Further evidence for MTL involvement in map-based navigation are provided by fMRI studies that explicitly compared strategy adoption. For example, using the virtual version of the MWMT (details in Subsection 1.3.2 and Figure 1.2), Parslow et al. (2004) found that hippocampal and parahippocampal activity increased during performance of task that required the participants to use map-based navigation, but not during the task that required route-based navigation (Parslow et al., 2004). Similarly, Shipman and Astur (2008) observed right hippocampal activity during the early stage of map-based navigation. This might reflect either: 1) an initial period of orienting oneself and determining the goal location, or 2) general memory encoding that is not necessarily specific for spatial navigation (Shipman and Astur, 2008). In a study by Jordan et al. (2004), participants were first asked to determine the shortest route between a start and an end position within a maze while being shown an overview of the maze. In a subsequent test phase, participants navigated between the start and end position in a virtual analog of the maze. In line with previous findings, compared to route-based navigation, map-based navigation was associated with increased activity in the left parahippocampal gyrus and the left hippocampus (Jordan et al., 2004; Zaehle et al., 2007).

Since map-based navigation relies on cognitive map, it is also important to understand how the MTL region supports the acquisition of survey knowledge. Early study by Grön et al. (2000) utilized a complex three dimensional maze containing several landmarks.

Participants needed to build and utilize survey knowledge as they searched for the way out of the maze. Analysis of the fMRI data, revealed significant activity in the right hippocampus and parahippocampal gyrus during navigation (Gron et al., 2000). Furthermore, Wolbers and Büchel (2005) used participants' improving knowledge of the spatial relationship between landmarks in a virtual town to assess the development of a survey knowledge. They found that the acquisition of new survey knowledge was associated with activity in the hippocampus (Wolbers and Büchel, 2005).

To encourage participants to develop a cognitive map of a virtual environment, Moffat et al. (2006) informed them of two subsequent test tasks that required processing of survey knowledge. These tasks included reproducing an aerial map of the environment, and determining the shortest route to a target object. The results showed significant activity in the MTL (Moffat et al., 2006). With a similar objective in mind, in a study by Iaria et al. (2007), participants freely explored a virtual city until they could correctly indicate the location of the available landmarks on an overview of the environment. This was done, during the encoding phase, to demonstrate the development of an accurate cognitive map. During a subsequent test phase, participants were asked to navigate between different pairs of landmarks via the shortest route. Unlike study by Moffat et al. (2006), analysis of the fMRI data revealed involvement of the hippocampal region in both the formation and the use of cognitive maps. The left anterior hippocampal activity was observed during the encoding phase. On the contrary, the right posterior hippocampal activity was observed during the test phase (Iaria et al., 2007). In line with all of these findings, more studies showed that the human hippocampus is associated with allocentric reference frame as well as the formation and the use of survey knowledge. This hippocampal-related functions allow accurate navigation from new starting locations based on either the configuration of environmental cues or the recognition of locations from a new viewpoint (Iaria et al., 2003; Doeller et al., 2008; Iglói et al., 2010; Lambrey et al., 2012).

Based on the findings presented above, it clear that fMRI remains one of the few methods for observing neuronal activity in deep brain structure such as the MTL. However, fMRI is not a direct measure of neuronal activity. Additionally, the exact relation between BOLD signal and the underlying neural activity remains an area of active research. A direct measure on how neuronal activity within MTL supports

navigation and wayfinding is obtained through invasive recordings. While patients with epilepsy performed a virtual navigation task, Ekstrom et al. (2003) recorded neural activity from 317 neurons. These neurons were located within the hippocampus, parahippocampal gyrus, as well as frontal lobes and amygdala. The navigation task consisted of freely exploring a two-dimensional VE while searching for passengers and delivering them to salient landmarks. The results showed that certain neuronal cells in the hippocampus robustly responded to spatial location (Ekstrom et al., 2003). In line with findings in rodents (Ferbinteanu and Shapiro, 2003; Ainge et al., 2007), many of these cells also showed modulations according to a patient's goal. In other words, the cells' firing rate changed based on which salient landmark a patient searched for. These cells, however, did not show changes in firing rather when a patient simply viewed the landmarks. In contrast, increased firing of the neuronal cells in parahippocampal gyrus was dependent on what landmark the patients viewed. This is in line with previous study in primate that noted view-responsive neurons within the parahippocampal gyrus (Rolls and O'Mara, 1995). Replicating the findings from Ekstrom et al. (2003), Jacobs et al. (2010) found place responsive neurons while epileptic patients navigated in a virtual circular environment. Similar to findings in rodent, the study found that these neurons tend to be directionally turned on the circular track. Simply put, they only fired when the patient navigated one way around the track and not the other (Jacobs et al., 2010). Taken together, these findings support a possible division of labor between hippocampus and parahippocampal gyrus. In such manner that neurons in the human hippocampus are responsive to place/location, and the one in parahippocampal gyrus are responsive to landmarks.

### **1.4.2 The Striatum Complex**

The striatum complex, which is located at the base of the forebrain, is a functional structure generally thought to be involved in the control of motor function (De Leonibus et al., 2001) and motivational processes (Delgado et al., 2008). Consequently, it has been commonly related to forms of cognitive process that requires the associations between instrumental response and a feedback (Kelley et al., 1997). More recent studies, however, demonstrate that manipulations of this region induce certain deficits in the processing of spatial information. For example pharmacological manipulations of the striatum have

been proven to affect performance in the Morris water maze, in the radial maze, or in tasks of spatial displacement (Colombo et al., 1989; Sutherland and Rodriguez, 1989; Roullet et al., 2001; Sargolini et al., 2003; Yin and Knowlton, 2004).

The striatal complex receives information from the whole cortical mantle. It is not a homogenous structure and thus, can be differentiated based on two characteristics. First, based on its intrinsic biochemical compartments (Sharp et al., 1986). Second, based on its connectivity, i.e. the diverse afferent/efferent projections (Alexander et al., 1990; Adams et al., 2001). With regard to spatial navigation and wayfinding, of particular interests are the dorsal striatum (dorsolateral and dorsomedial) and nucleus accumbens. These two structures are thought to support stimulus-response associations and procedural memories, which are crucial for route-based navigation (De Leonibus et al., 2005).

The caudate nucleus plays an important role in learning and spatial memory. It is often activated during virtual navigation tasks in functional imaging studies (Iaria et al., 2003; Maguire et al. 1998; Moffat et al., 2006). Because of its extensive connections with the prefrontal cortex and hippocampus, the caudate is a part of distributed fronto-striatal and striato-hippocampal system. Consequently, studies investigating contributions of the caudate to human spatial navigation suggest that it may work in concert with the hippocampal systems (Voermans et al. 2004). In addition, it may also play a role in non-spatial or procedural response (Hartley et al. 2003).

Furthermore, fronto-striatal and striato-hippocampal system are especially vulnerable to aging. Therefore, it is not surprising that in older adults, the caudate nucleus plays a critical role in supporting navigation, sometimes even more than the hippocampus or other brain areas within the MTL. Study by Moffat et al. (2007) found that larger caudate volume was associated with better performance in both older subjects.

### **1.4.3 The Retrosplenial Complex (RSC)**

This subsection focuses on the functional region of retrosplenial cortex/posterior cingulate/medial parietal region, close to the point where the calcarine sulcus joins the

parietal-occipital sulcus (Vann et al., 2009). The central location of this functional region makes it pivotally positioned to receive information from, and readily influence, many key brain regions responsible for the processing of spatial information. As retrosplenial cortex (BA 29 and 30) adjoins and is partially encircled by the posterior cingulate (BA 23 and 31), these labels are often used somewhat interchangeably when describing the locations of functional activations (Van Hoesen and Pandya, 1975). Because of this ambiguities, this doctoral thesis uses the term retrosplenial complex (RSC, Figure 1.8) to refer to this functionally-defined region, which is not necessarily identical to the anatomically-defined retrosplenial cortex.

Although there is clear evidence from both human and animal studies pinpoints RSC role in spatial cognition, there is no consensus as to its precise correlate when it comes to the dichotomy of route-based and map-based navigation. Instead, findings from lesion, single neuron, and neuroimaging studies propose that RSC functions in spatial navigation and wayfinding fall into three categories: (1) landmark processing, (2) spatial references, (3) spatial schema. The first category, landmark processing, argues that RSC has a specific function in the encoding of the spatial and directional characteristics of landmarks, independent of their identity. Second, with regard to spatial references, studies reported that RSC mediate the translation of information between different spatial references. Thus, allowing one to translate *'you are here information'* into *'your destination is to the left information'*. The last category, which is by far the broadest of all, proposes that RSC is involved in formation and consolidation of hippocampus-dependent spatial as well as episodic memories. Based on this three categories of RSC functions, it is not surprising that this region has been implicated in both route and map-based navigation.

One of the earliest evidence for RSC's role in landmark processing comes from reported navigational difficulties when the regions is damaged either by stroke in humans or by lesions in rodents (Vann et al., 2009). Patients with retrosplenial damages have been reported to recognize buildings and landscape. However, the landmarks did not provoke directional information. Consequently, these patients could not determine which direction to proceed (Bottini et al., 1990; Takahashi and Kawamura, 2002; Greene et al., 2006; Ino et al., 2007; Osawa et al., 2008). With regard to RSC's role in integrating different spatial reference frames, patients with RSC damage could not use map to

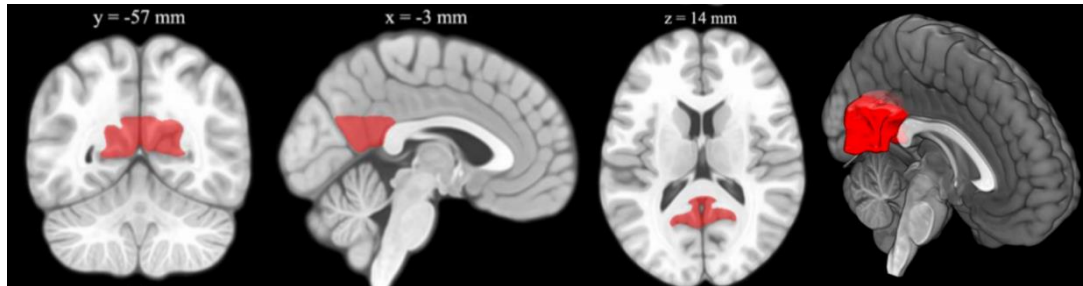


indicate view point relative to certain landmark (Katayama et al., 1999). In line with these findings, other lesion studies reported patients' difficulty in following a route when change of orientation is required (Miller et al., 2014). This inability to convert back and forth between reference frames denotes the underlying feature of RSC such that it has access to the same spatial information represented in different ways, and is needed in order to switch between egocentric and allocentric representations.

Single neuron studies in rodent indicates that RSC neurons can encode a variety of spatial quantities including that of head direction (HD) cells (Chen et al., 1994a; Chen et al., 1994b). These are cells that fire preferentially when animal faces in a particular global direction (Taube, 2007). A later study by Jacob et al. (2017) found a subpopulation of HD cells, the firing of which was controlled by the local environmental cues independently of the global HD signal (Jacob et al., 2017). In 2015 and later on in 2017, Alexander and Nitz found RSC cells that respond to specific combinations of location, direction, and movement (Alexander and Nitz, 2015, 2017). In addition to electrophysiological recordings in rodents, recordings from medial parietal neurons of macaque monkeys advocate the role of RSC in relating local and global spatial reference frames (Sato et al., 2006). Of the many neurons that responded strongly during the task, which involved following well-learned routes through a multi-room virtual-reality environment.

Since the advent of fMRI in cognitive neuroscience, many studies have investigated RSC activation as human subjects perform tasks in the scanner, as described in Section 1.3. One of the earliest fMRI study show that RSC activation occurred during scene viewing and imagery (Park and Chun, 2009), mental imagination of navigation through familiar environment (Epstein and Higgins, 2007), route learning (Wolbers and Büchel, 2005), spontaneous trajectory changes as well as confirmation of expectations about the upcoming features of the outside environments. Other fMRI studies confirmed that RSC activity was specifically associated with thought of location and orientation, as opposed to simple object recognition (Epstein and Higgins, 2007; Iaria et al., 2007). Both studies also show that pattern of RSC activation differs from that of hippocampus, such that RSC is strongly active during both encoding and retrieval of spatial information. Similar line of works have investigated the encoding of location and/or direction as well as local or individual landmarks by RSC (Marchette et al., 2014; Robertson et al., 2016). In

contrast to these two studies that focus on RSC's role in local encoding of direction, study by Shine et al. (2016) find evidence for global heading representation in the RSC. Given that there is evidence for both local and global encoding of directions in the RSC, it is often argued that RSC is involved in processes that require adaptation of one or more navigation strategies. This is especially true when environmental cues are being used.



**Figure 1.8. Schematic of the Retrosplenial Complex (RSC)**

Adapted from Burles et al. (2017)

## 1.5 Value-based Decision Making

Value-based decision making is defined as choosing among actions based on their relative value of potential consequences in order to maximize reward and minimize punishment (Rangel et al., 2008). Similar to spatial navigation, there are at least two complimentary strategies involve in value-based decision making namely: model-free choice and model-based choice. Research in decision neuroscience has been seeking to build a biologically sound theory of how humans and animals employ these two choice strategies. The center of this effort is combining computational models and neuroimaging methods, such as single unit recordings and fMRI, to identify the types of signals and signal dynamics that are required by different value-dependent decision problems (O'Doherty et al., 2003b; O'Doherty et al., 2004; Daw et al., 2006; Valentin et al., 2007; Wunderlich et al., 2009; Glascher et al., 2010).

Starting from this section, this chapter will focus on reviewing a computational framework from value-based decision making that is useful for dissecting different cognitive processes behind strategy adoption in spatial navigation and wayfinding. The review has three components. First, this section focuses on decision processes and how the two choice strategies support these processes. Second, Section 1.6 focuses on reinforcement learning (RL) algorithms that formally provides a computational framework in modeling choice strategies. Lastly, Section 1.7 focuses on neural correlates of reinforcement learning's key variables in value-based decision making.

### 1.5.1 Cognitive Processes in Value-based Decision Making

There are five basic cognitive processes that are required for value-based decision making: identification, valuation, action selection, outcome evaluation, and updating. To better illustrate these processes, the explanation in this as well as the following subsection is organized around the diagram in Figure 1.9 and the simple rodent maze task shown in Figure 1.10. The maze has three decision points (A, B, and C) and four possible outcomes (cheese, nothing, water, and carrots). At each decision point, there are two possible actions: going left or going right.

The first cognitive process in value-based decision making is the identification of decision problem, which entails identifying internal states, external states, and potential course of action. In the simple maze example, there are three internal states: level of hunger, level of thirst, and level of cheese craving. In the subsequent cognitive process, the different actions that are under consideration need to be assigned a value based on the internal and external states. In order to make appropriate decisions, these value have to be a reliable predictors of the benefits that are likely results from each action. Going back to the maze example, when the rat is hungry, the cheese is assigned the highest value. Third, the different values need to be compared for the rat to be able to make a decision, i.e. select an action. When the rat is hungry, the cheese is more valuable than the water or carrot. In contrast, when the rat is thirsty, the order of valuation becomes water, carrot, and cheese. Fourth, after implementing the decision, the rat needs to measure the desirability of the outcome. This cognitive process is referred as outcome evaluation. Lastly, the feedback measures are used to update the other processes to improve the quality of future decisions.

It is important to note that these five cognitive processes are not rigid. For instance, it is still debatable whether valuation (process no. 2) must occur before action selection (process no. 3), or whether both computations are performed in parallel. Nevertheless, the categorization of these processes is conceptually useful to breakdown the value-based decision into a testable constituent processes.

Furthermore, on the basis of a sizable body of animal and human behavioral evidence, there are at least two strategies to assign values during decision making processes: model-free choice and model-based choice. Subsection 1.5.2 and 1.5.3 review the operational division of the two valuation strategies according to the style of computations that is performed by each.

### **1.5.2 Valuation Based on Model-Free Choice**

Model-free choice can learn, through repeated training, to assign value to a large number of actions (Dayan and Niv, 2008). The valuation based on model-free choice have four key characteristics. First, values are assigned to stimulus-response associations, which

indicate the action that should be taken in a particular state of the world, on the basis of previous experience. This is done through a series of trial and error without constructing a model or representation of the environment.

Second, subject to some technical qualifications, model-free choice assign values to actions that is commensurate with the expected outcome that these action generate. This holds true as long sufficient practice is provided and the environment is sufficiently stable. Going back to the maze example, Figure 1.10. C shows the valuation based on model-free choice, i.e. model-free values, of each action at each location in the maze, assuming that the rat chooses optimally for the internal state of hunger. Such model-free values can be used without direct reference to a model of transitions or outcomes (Rangel et al., 2008).

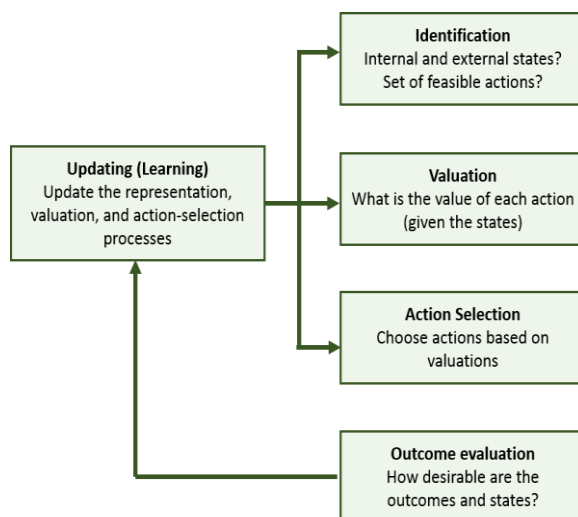
Third, because values are learned by trial and error, model-free choice is believed to be relatively slow. Consequently, it might forecast the value of actions incorrectly immediately after a change in the action-reward contingencies. Lastly, model-free choice is inflexible. Model-free values, such as those shown in Figure 1.10. C, are just numbers, unrelated from the outcomes that underlie them, or the statistics of the transitions in the environment. Should there be a change in the internal state, the model-free values will not change without further and statistically expensive updating (Rangel et al., 2008).

### **1.5.3 Valuation Based on Model-Based Choice**

In contrast to valuation based on model-free choice, model-based choice assign values using a model, or representation of state transition, of the environment. This representation should indicate the probability with which decision progress from one decision point to the next. The progression of the decision is based on which action is taken and the likely outcomes at each decision point (Daw et al., 2011). Figure 1.10. D shows the model of the simple maze. This representation is nothing more than the three of locations in the maze (external states) joined according to the actions that lead between them.

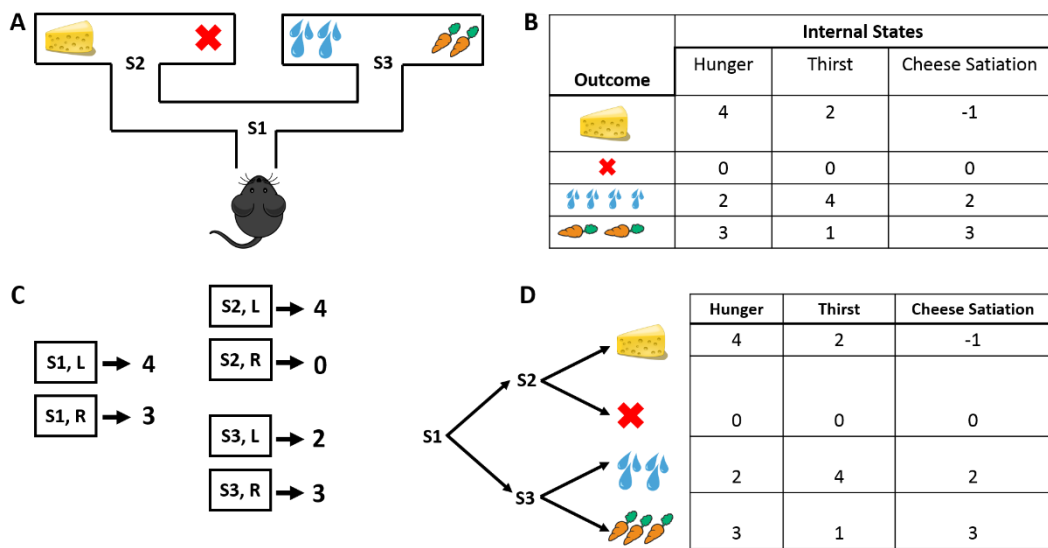
Given some systematic way to select action at each external state, models such as that shown in Figure 1.10. D, admit a conceptually very simple way of making predictions about the values of states or locations in the maze, namely searching forward in the model, accumulating expected values all the while. Unfortunately, valuation based on model-based choice, i.e. model-based values, places a huge burden on working memory. Thus, it is not as efficient as model-free approach.

Note that an important difference between model-free choice and model-based choice has to do with how they respond to changes in the environment. This is because to carry out the necessary value computation, the model-based choice need to store both action-outcome and outcome-value association. Consider the maze example and the valuations made by the rat, which has learned to obtain cheese, after it consumed the cheese to the level of satiation. The model-based choice has learned to associate the action ‘go left’ at decision point ‘B’ with the outcome ‘cheese’ and thus assign a value to the [B, go left] pair that is equal to the current value of ‘cheese’. In this example the value ‘cheese’ is low because the rat has consumed it to satiation. In contrast, the model-free choice will still assign a high value to the [B, go left] pair because this is the value that the rat learned during the pre-satiation time.



**Figure 1.9. Processes in the value-based decision making**

Value-based decision making can be broken down into five basic processes. First, the construction of the decision problem. This entails identifying internal and external states as well as potential course of action. Second, the valuation of different actions under consideration. Third, the selection of one of the actions on the basis of their valuations. Fourth, after implementing the decision the brain needs to measure the desirability of the outcomes that follow. Finally, the outcome evaluation is used to update the other processes to improve the quality of future decisions.



**Figure 1.10. Model-based and model-free choice in a simplified maze task**

*A.* A simple maze with three external states (S1, S2, S3), three internal states (hunger, thirst, and cheese satiation), two possible actions at each external state (left or right), and four outcomes (cheese, nothing, water, and carrot).

*B.* The values of the outcomes under different internal states.

*C.* Valuation based on model-free choice. Immediately after cheese satiation, these values do not change. It is only after direct experience with the cheese that the value associated with (S2, L) and subsequently (S1,L) is reduced.

*D.* A tree-based model of the state action environment, which can be used to guide decision at each state and assign values according to the model-based choice.

## 1.6 Reinforcement Learning as a Computational Framework to Represent Choice Behavior

Reinforcement learning (RL) was originally born out of mathematical psychology and operation research. It provides both qualitative and quantitative computational models on how natural and artificial systems can learn to predict the consequence of and optimize their behavior in making value-based decisions within certain environments (Sutton and Barto, 1998; Dayan and Niv, 2008). These decision environments are characterized by a few key concepts as illustrated in Figure 1.11. These key concepts are agent, environment, state spaces, sets of actions, a reward signal, a policy, a value function, and optionally a model of the environment.

First, an agent is simply the decision maker. Everything outside the agent, with which it interacts, is called the environment. A state space or simply *state* is a signal conveying to the agent some sense of the condition of the environment at a particular time. In a more general sense, a state is defined as whatever information is available to the agent about its environment. A state space can take the form of board positions in a game, the existence or absence of different stimuli in an operant box, or, in the case of spatial navigation, a state space can simply be a specific location in a maze or in a city. An agent interacts with its environment by choosing one of several available actions. The agent then receives feedback, or reinforcement, from the decision environment in the form of a reward signal. A reward signal is defined as an immediate, possibly stochastic payoff that results from performing an action in a state. The learning task for the agent is to optimize either a sum or an average value of future rewards. A policy is how the learning agent maps perceived states of the environment to actions, which are about to be taken in those states. A value function is an estimate of the total, possibly discounted, future reward. A value function is computed to improve the policy. In other words, a value function indicates what is desirable in the long run. Lastly, for the environment within which an agent takes actions, the agent can build a transition model to plan actions. This is done by considering possible future states even before those states are actually experienced.

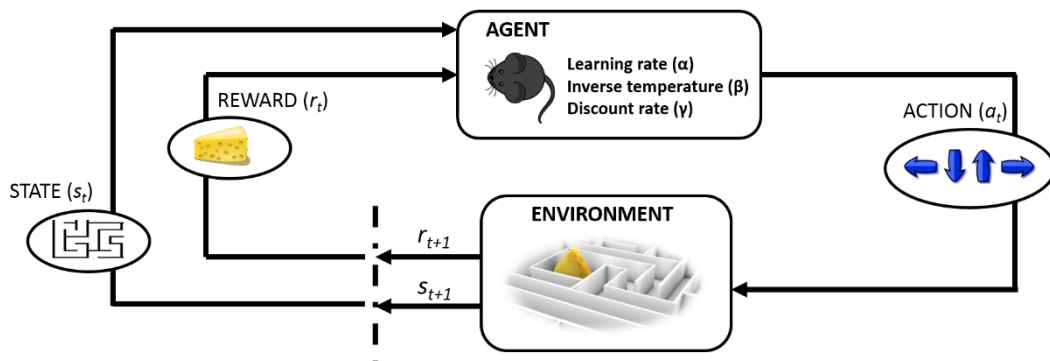
RL algorithms can be divided into two broad classes, model-based and model-free, which perform optimization of choices in different ways. In this section, different



model-free and model-based RL algorithms are reviewed. Of particular interests are those that are used as computational framework to explain different valuation strategies in value-based decision making.

### 1.6.1 Model-free Reinforcement Learning

Model-free reinforcement learning (RL) uses experience to learn directly one or more simpler quantities (state values, action values, state/action values, or policies). Thus, model-free RL can achieve optimal behavior without constructing a representation of the environment. For a certain policy, a state has a value, which is defined in terms of the future utility that is expected to accumulate starting from that state. Crucially, correct values satisfy a set of mutual consistency conditions such that a state can have high value only if the actions specified at that state lead to one of these two conditions: (1) immediate outcomes with high utilities, and/or (2) states which promise large future expected utilities. Model free algorithms, in contrast to their model-based counterparts, are statistically less efficient. This is because the information from the environment is combined with previous, and possibly erroneous, estimates or beliefs about state values, rather than being used directly. Examples of model-free algorithms that are used to represent model-free choices in value-based decision making include, but not limited to Temporal Difference (TD) learning, Q-Learning, and SARSA ( $\lambda$ ).



**Figure 1.11. The agent-environment interface and elements of reinforcement learning**

An agent interacts with its environment by selecting certain actions. The environment provides feedback to the agent through a reward signal, which is defined as the goal in a reinforcement learning problem. On each time step, the environment sends to the reinforcement learning agent a single number, i.e. the reward. The agent's sole objective is to maximize the total reward it receives over the long run.

### ***Temporal Difference Learning***

The most straightforward approach to learning state values is by sampling. This simply means that the value of the state is updated given the reward the agent receives such that the value is equivalent to the immediate expected reward  $V_s = r$ . This is known as Rescola-Wagner learning algorithm. The basic idea behind this algorithm is that learning should occur only when observed events violate expectations. Thus, the value  $V$  of state  $S$  is updated with a prediction error:

$$\delta \leftarrow r - V_s$$

and

$$V_s \leftarrow V_s + \alpha \delta$$

where  $\alpha$  is the learning rate. This learning rate determines to what extent the newly acquired information will override the old information. A factor of 0 means that the agent does not learn anything, while a factor of 1 would make the agent consider only the most recent information.

An expansion of the Rescola-Wagner learning algorithm is the Temporal Difference (TD) learning. The TD learning algorithm divides a series of action selection into smaller time points. At each time point  $t$ , an agent experiences a state  $s_t$ , which produces a reward  $r_t$ . The goal of the agent is to estimate the value of a state  $V(s_t)$  in terms of its cumulative future rewards. The prediction error is then replaced by:

$$\delta_t \leftarrow r_t + \gamma V(s_{t+1}) - V(s_t),$$

which is called TD prediction error and the value is updated using the following equation:

$$V_{k+1}(s_t) \leftarrow V_k(s_t) + \alpha \delta_t.$$

Unlike the Rescola-Wagner algorithm, TD considers not only the immediate reward  $r_t$ , but also accounts for the sum over all the rewards in the subsequent states. This way, TD algorithms extend the discrete Rescola-Wagner algorithm into a continuous time learning (O'Doherty et al., 2003b; Parslow et al., 2004; Samejima et al., 2005).

### ***Q Learning***

Another commonly used model-free RL algorithm is the  $Q$ -learning, which explicitly learns the state-action value  $Q(s,a)$ . The way in which the state action value is updated is a variation on the theme of temporal difference learning and formulated as follow:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \delta_t,$$

where  $\delta_t$  is the TD reward prediction error and is computed by:

$$\delta_t \leftarrow r_t + \gamma \underbrace{\max_a}_{a} Q(s_{t+1}, a_t) - Q(s_t, a_t).$$

The update takes place when the agent choose an action  $a_t$ , and receives a reward  $r_t$  by moving from states  $s_t$  to  $s_{t+1}$ . In general, the Q-learning algorithm will converge to an optimal policy if the learning rate  $\alpha$  decreases properly and all the state-action pairs are visited infinitely often (Tanaka et al., 2004; Knutson and Cooper, 2005; Daw et al., 2006; Glascher et al., 2010).

### ***SARSA (Q): State-Action-Reward-State-Action with Eligibility Trace***

The name SARSA reflects the fact that the main function for updating the state action value  $Q(s,a)$  depends not only on the current state of the agent  $s_t$ , the chosen action  $a_t$ , the reward the agent gets for choosing this action, but also  $s_{t+1}$  the state the agent will now be after taking action at as well as the next action  $a_{t+1}$  the agent will choose in its new state. This approach is formulated as follow:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)].$$

A SARSA agent will interact with the environment and update the policy based on actions taken (instead of the minimum or maximum  $Q$ -value of the next state), known as an *on-policy* algorithm. In other words, it updates the  $Q$ -value of the policy being executed instead of the optimal  $Q$ -value. As expressed above, the  $Q(s,a)$  value for a state-action is updated by an error, adjusted by the learning rate alpha.  $Q(s,a)$  values represent the possible reward received in the next time step for taking action  $a_t$  in state  $s_t$ , plus the discounted future reward ( $r_{t+1}$ ) received from the next state-action observation.

One disadvantage of SARSA algorithm is that in a situation where a series of actions, instead of a single action, is needed to reach reward, SARSA only updates the  $Q$ -value of the last action in the sequence. In many applications, such as in spatial navigation, it is desirable to be able to update a series of actions very fast and to be able to take into account anything that has changed. A method in which the update of  $Q$ -value extends over  $n$  steps within a series of actions is called  $n$ -step SARSA.

Furthermore, an extension of  $n$ -step SARSA is implementing eligibility traces to SARSA algorithm, known as SARSA ( $\lambda$ ). An eligibility trace  $\lambda$  is a temporary record of the occurrence of an event, such as visiting a state or the taking of an action. The trace marks the memory parameters associated with the event as eligible for undergoing learning changes. Using this parameter, the algorithm can distribute reward or penalty into previously taken series of actions more efficiently. Thus, unlike the  $n$ -step method that updates the values of actions within a sequence equally, eligibility trace ensures that the effect of older actions or visited states is less pronounced for the resulting behavior (Daw et al., 2006; Glascher et al., 2010; Simon and Daw., 2011). The complete SASRA ( $\lambda$ ) algorithm is presented in Figure 1.12. Figure 1.13 illustrates the difference between one-step SARSA,  $n$ -step SARSA, and SARSA ( $\lambda$ ).

**SARSA ( $\lambda$ ): State-Action-Reward-State-Action with Eligibility Trace**

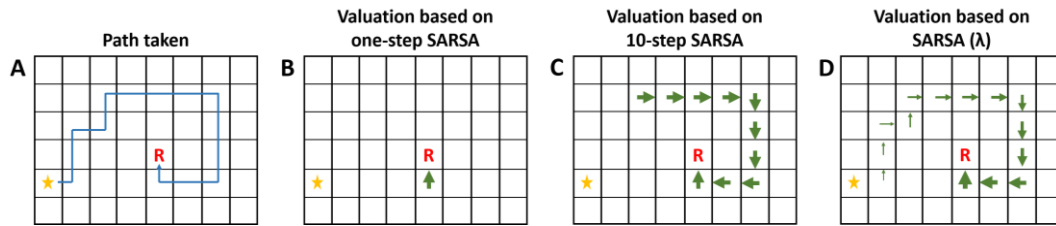
```

Initialize  $Q(s,a)$  arbitrarily
Repeat (for each episode):
   $e(s_t, a_t) \leftarrow 0$ , for all state action pair
  Initialize  $s, a$ 
  Repeat (for each step of the episode):
    Take action  $a_t$ , observe  $r_{t+1}, s_{t+1}$ 
    Choose  $a_{t+1}$  from  $s_{t+1}$  using policy derived from  $Q$ 
     $\delta \leftarrow r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)$ 
     $e(s_t, a_t) \leftarrow e(s_t, a_t) + 1$ 
    For all state action pair:
       $Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \delta e(s_t, a_t)$ 
       $e(s_t, a_t) \leftarrow \gamma \lambda e(s_t, a_t)$ 
     $s_t \leftarrow s_{t+1}; a_t \leftarrow a_{t+1}$ 
  until  $s$  is terminal

```

**Figure 1.12. The SARSA with eligibility trace algorithm**

The  $Q$ -value is updated at every step and weighted according to the geometrical distribution in every episode. The trace decay parameter  $\lambda$  for representing the use of eligibility traces in the algorithm.



**Figure 1.13. Comparison of valuation based on various SARSA algorithms**

*A.* The path taken by an agent in a single episode. The initial estimated values were zeros except for a positive reward at the location marked by **R**. The arrows in the other panel show, for various SARSA algorithms, which action-values would be increased, and by how much upon reaching reward **R**.

*B.* One-step SARSA would only update the last action value, i.e. the one directly lead to **R**.

*C.*  $n$ -step SARSA (in this case  $n = 10$ ) equally update  $n$  action values.

*D.* SARSA with eligibility trace updates all the action values up to the beginning of the episode to different degrees, fading with recency. This algorithm provides the best tradeoff, strongly learning how to reach the goal from the right, yet not as strongly learning the roundabout path to the reward from the left that was taken in this single episode.

## 1.6.2 Model-based Reinforcement Learning

Model-based RL refers to learning optimal behavior indirectly by learning a model of the environment by taking actions and observing the outcomes that include the next state and the immediate reward (Wunderlich et al., 2011; Glascher et al., 2011; Glascher et al., 2012; Kahn et al., 2014). By a model of the environment, we are referring to anything that an agent can use to predict how the environment will respond to its actions (Dayan and Niv, 2008; Rangel et al., 2008). The models are used in lieu of or in addition to interaction with the environment to learn optimal policies. Moreover, given a state and an action, a model produces a prediction of the resultant next state and next reward. If the model is stochastic, then there are several possible next states and next rewards, each with some probability of occurring.

Given a complete model of the environment, there is a collection of algorithms that can be used to compute values. This collection of algorithms is called Dynamic Programming (DP). The key idea of DP is that optimal policies can be obtained once optimal value functions satisfy the following Bellman optimality equations:

$$V^*(s) = \underbrace{\max}_a E \langle R_{t+1} + \gamma V^*(S_{t+1}) | S_t = s, A_t = a \rangle$$

$$V^*(s) = \underbrace{\max}_a \sum_r^{s'} p(s', r | s, a) [r + \gamma V^*(s')] ]$$

or

$$Q^*(s, a) = E \left\langle R_{t+1} + \gamma \underbrace{\max}_a Q^*(S_{t+1}, a') \middle| S_t = s, A_t = a \right\rangle$$

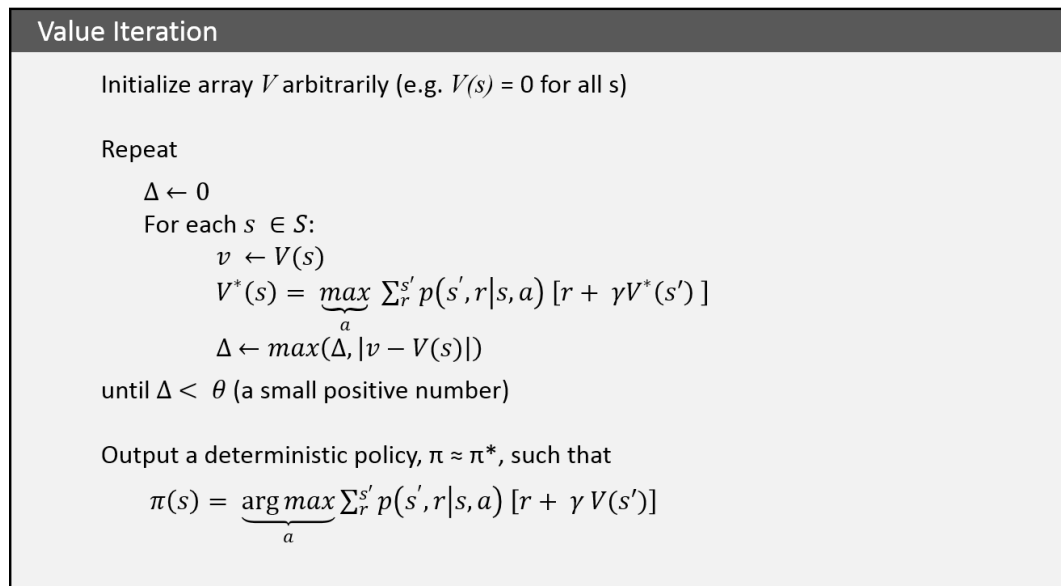
$$Q^*(s, a) = \sum_r^{s'} p(s', r | s, a) \left[ r + \gamma \underbrace{\max}_a Q^*(s', a') \right]$$

### **Value Iteration**

Common DP algorithms used to represent model-based choices are value iteration and policy iteration. In these approaches, both algorithms start with a randomly initialized value function or policy. In policy iteration, the algorithm calculates the value of the current policy and then loops through the state space, updating the current policy to be greedy with respect to the backed up values. This is repeated until the policy converges.

In value iteration, the algorithm loops through the state space, updating the value estimates of each state using Bellman backups, until convergence. In other words, value iteration starts at the *end* and then works backward, refining an estimate of either  $Q^*$  or  $V^*$ .

Value iteration formally requires an infinite number of iterations to converge exactly to  $v^*$ . In practice, iteration stops once the value function changes by only a small amount in a sweep. Figure 1.14 shows a complete algorithm with this kind of termination condition.



**Figure 1.14. Value iteration algorithm**

Model-based reinforcement learning based on dynamic programming.

### 1.6.3 Action Selection

One specific challenge in RL, which is not encountered in other types of learning, is the tradeoff between exploration and exploitation, i.e. explore-exploit dilemma. In order to accumulate rewards, an agent must prefer actions that it has tried in the past and found to be effective in producing reward, i.e. exploit. To discover such actions, however, an agent has to try actions that it has not chosen before, i.e. explore.

An effective and popular means of balancing exploitation and exploration in reinforcement learning is by the mean of  $\epsilon$ -greedy. Epsilon greedy ( $\epsilon$ -greedy) is a way of

selecting random actions with uniform distribution from a set of available actions. Using this approach, a reinforcement learning agent can select either random action with  $\epsilon$  probability or an action that gives maximum reward in a given state with  $1 - \epsilon$  probability. For instance, if in a given state an agent has to make 100 decisions, this method select random actions in 20 decisions if the value of  $\epsilon$  is 0.2. In other words, the agent exploit in 80 decisions while explore in 20 decisions.

One drawback of  $\epsilon$ -greedy is that when it explores, it chooses equally among all actions. This means that it is just as likely to choose the worst appearing action as it is to choose the next-to-best. To overcome this drawback, the obvious solution is to vary the action selection probabilities as a graded function of estimated value. With this approach, the greedy action is still given the highest selection probability. All other actions are ranked and weighted according to their value estimates. This approach is called *softmax* action selection rules, which happens to be the most widely used approach in the studies of value-based decision making (Knutson and Cooper, 2005; Preuschoff et al., 2006; Hare et al., 2008; Preuschoff et al., 2008; Daw et al., 2011). The most common softmax method, which is the one used in this thesis, uses a Gibbs or Boltzmann distribution. The probability of choosing action  $a$  given state  $s$ , is given by the following equation.

$$Pr\{A_t = a|s\} = \frac{e^{Q(s,a)\beta}}{\sum_{b=1}^k e^{Q(s,b)\beta}}$$

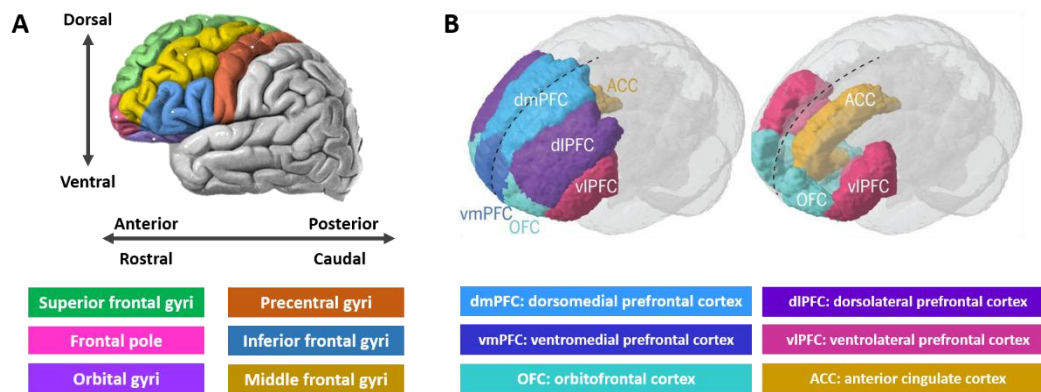
Parameter  $\beta$  represents inverse temperature. For high temperatures, all actions have nearly the same probability. The lower the temperature, the more expected rewards affect the probability of choosing certain action.



## **1.7 Neural Correlates of Reinforcement Learning during Value-based Decision Making**

This doctoral thesis uses fMRI to test the neural validity of key internal variables from reinforcement learning models in predicting choice behavior during route-based and map-based navigation. This method, combining reinforcement learning models with fMRI data, is a specific fMRI approach that estimates how changes in the BOLD signal correlate with quantitative computational predictions of neural activity (Dayan and Niv, 2008). Thus, it allows hidden variables and computational processes to be uncovered in ways not possible with traditional event related or parametric paradigm design (O'Doherty et al., 2003b; Parslow et al., 2004; Samejima et al., 2005; Daw et al., 2006; Hampton et al., 2006; Plassmann et al., 2007; Valentin et al., 2007; Hare et al., 2008; FitzGerald et al., 2009; Wunderlich et al., 2009; Daw et al., 2011; Simon and Daw, 2011; Wunderlich et al., 2011; Hunt et al., 2012).

Using the same approach, along with findings from single unit recordings in rodents and primates, neural signals in different brain regions have been found to correlate with reward expectations, valuation, and outcome evaluation. Among those regions, the network of prefrontal and striatal regions remains the most well-documented findings in decision neuroscience. The questions of which reinforcement learning algorithms are best represented at neuronal level, and specifically, what brain structures execute different key variables of those algorithms are the primary focus of this section.



**Figure 1.15. Major subdivisions of the frontal lobe**

A. Lateral view of major anatomical subdivisions of the frontal lobe. Adapted from Frank Gaillard, Radiopedia.org, rID: 46670.

B. Major functional subdivisions of the prefrontal cortex (PFC). Adapted from Carlen (2017).

### 1.7.1 Prefrontal Cortex

As illustrated in Figure 1.15, the prefrontal cortex is the most anterior part of the frontal lobes of the brain. The posterior part of the frontal lobe forms the motor and premotor areas. Specifically elaborated in primates and humans, it is the cortical region that underwent the greatest expansion during evolution. Due to a late myelination of the axonal connections, the peak of its maturity arrives only at the end of adolescence (Wilson et al., 2010). Connectivity patterns to this area highlight a distinction between an orbitofrontal, a medial, and a lateral sub-networks in the PFC, a distinction that has been shown to have functional implications especially with regards to value-based decision making (Lee et al., 2012). For the remainder of this subsection, neural correlates of value-based decision making within the PFC is discussed based on the following functional subdivisions of the PFC: orbitofrontal cortex (OFC) and ventromedial prefrontal cortex (vmPFC), dorsolateral prefrontal cortex (dlPFC), and anterior cingulate cortex (ACC).

#### ***Orbitofrontal and Ventromedial Prefrontal Cortex***

The orbitofrontal cortex (OFC) is a large are cortical area located at the most ventral surface of the prefrontal cortex. It is situated directly above the orbit of the eye, hence the name, and includes parts of the medial wall between the hemispheres. The OFC consists of BA 11 and 47. A noteworthy feature of OFC's anatomy is its connectivity. It

has remarkably close connections to all sensory areas as well as widespread connections to other parts of the frontal cortex, striatum, amygdala, and the hippocampus.

While the ventromedial prefrontal cortex (vmPFC) does not have a universally agreed on demarcation, in this thesis the functional subdivision follow the ventromedial reward network of Ongur and Price (2000). This network includes BA 10, 14, 25, and 32. It is connected to and receives input from the ventral tegmental area, amygdala, the temporal lobe, the olfactory system, and the dorsomedial thalamus. In turn, it sends signal to many different brain regions including the temporal lobe, the hippocampus, and the cingulate cortex (Ongur and Price, 2000).

Findings from lesion, single-unit recordings, and neuroimaging studies propose that vmPFC and OFC functions in value-based decision making fall into two categories: (1) global currency integrator, and (2) estimation of the expected value of each stimulus based on past experience.

As global currency integrator, OFC represents economic value associated with a goal, in particular value that has to be calculated on-the-fly rather than learned from experience. This notion is supported by various recordings of neural activity in the OFC of monkeys (Tremblay and Schultz, 1999; Padoa-Schioppa and Assad, 2006). In particular, in the study by Padoa-Schioppa and Assad (2006), the authors showed that a proportion of recorded neurons showed firing activity that varies linearly with the subjective value of the goal regardless of the chosen option.

This general finding, a reward representation that is independent of sensory or motor aspects of the option, is supported by a number of neuroimaging studies in humans. It is important to note that brain activity in the OFC and vmPFC, similar to findings in monkeys, does not only respond only to money. Instead, these areas are global currency integrators. Kim et al. (2006) reported an increase in activity in the vmPFC after successful avoidance of punishment. This avoidance of a negative reward can be seen as an intrinsic reward correlation, serving to reinforce avoidance (Kim et al., 2006). Moreover, Plassmann et al. (2007) showed correlations of the willingness-to-pay in the vmPFC. This willingness-to-pay, the maximum amount of money subjects are willing to pay in exchange for the presented choices, represents the value of a goal (Plassmann et

al., 2007). Again, this pattern has been shown in variety of goals whether it is pleasant odor (Rolls et al., 2003; Grabenhorst et al., 2007), tastes (Small et al., 2003), or face attractiveness (O'Doherty et al., 2003a).

Alternatively instead of solely monitoring goal values and subsequent outcomes, expected values of different choices can be first estimated and then used to identify the choice with the largest benefit. Electrophysiological studies support this notion. In both studies by Padoa-Schioppa and Assad (2006 and 2008), some cells in the vmPFC are correlating positively with the chosen value, whereas others correlate negatively.

Through neuroimaging findings, global correlation of estimated values can be observed. Values derived from reinforcement learning models are positively correlated with activity in the vmPFC (Tanaka et al., 2004; Knutson and Cooper, 2005; Daw et al., 2006; Glascher et al., 2010). This correlation is even observed for expected values of delayed rewards (Glimcher and Rustichini, 2004). Besides these typical results, some studies also reported that brain activity in the vmPFC reflects expectations of monetary losses (Tom et al., 2007; Basten et al., 2010). Boorman et al. (2009) demonstrated that vmPFC correlated positively with the chosen option and negatively with the rejected one, signaling a comparison between the different options (Boorman et al., 2009).

### ***Dorsolateral Prefrontal Cortex***

The dorsolateral prefrontal cortex (dlPFC) serves as the end point of for the dorsal pathway. It is the most recently evolved and last to develop in adulthood. This part of the brain does not receive sensory information for the visual or auditory areas. It does not send motor command to the muscle. Instead, it is the association cortex that bridges input and output. Consequently, it is mainly involved with pausing and choosing action as well as planning and action sequencing.

From seminal neurological observations to modern functional neuroimaging, substantial evidence implicates the dlPFC in model-based choice system. Findings from fMRI studies show that abstract action representations, such as plans for simple sequences or rules for selecting action based on context, are processed in the dlPFC (Petrides and Pandya, 1999). Moreover, to carry out the necessary computations, model-based choice needs to store the action-outcome and outcome-values associations. Studies across

methods and species showed that dlPFC encode appetitive goal-value signals in monkeys (Wallis and Miller, 2003; Barraclough et al., 2004) as well as humans (Plassmann et al., 2007).

### ***Anterior Cingulate Cortex***

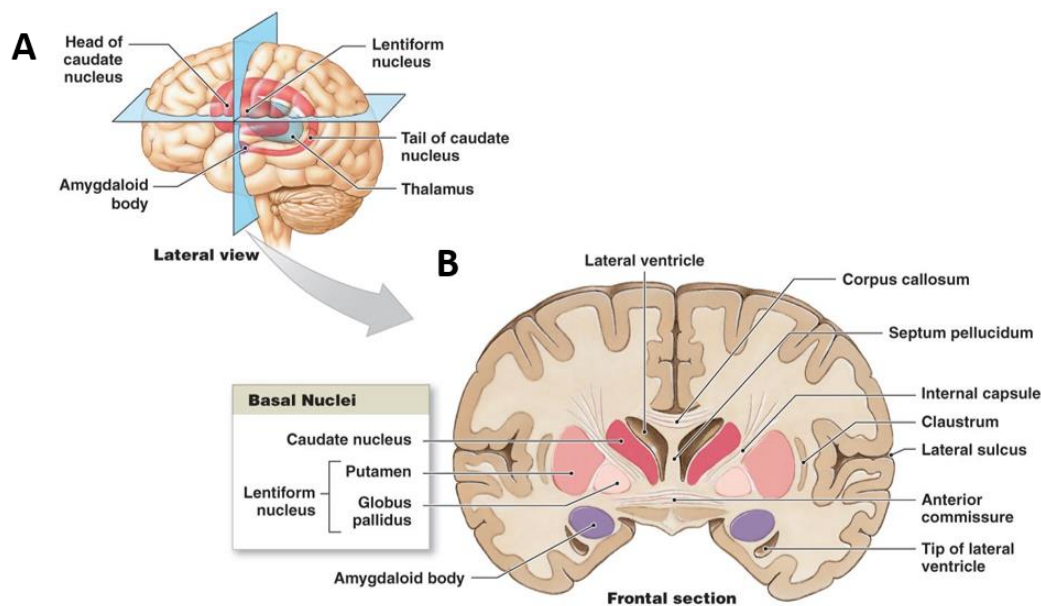
The cingulate cortex lies in the medial wall of each cerebral hemisphere, above and adjacent to the corpus callosum. The more anterior portion surrounding the genu of the corpus callosum has been referred as rostral or ventral anterior cingulate cortex (ACC) and consists of BA 24, 25, 32, and 33. This area has connections to both the *emotional* limbic system and *cognitive* prefrontal cortex. Thus, the ACC has an important role for affect regulation as well as value-based decision making (Stevens et al., 2011).

The roles of ACC in value-based decision making are well reported in both fMRI experiment on humans and single-unit recordings in monkeys. In general, these roles can be divided in two broad categories: (1) interaction of choice decisions and consequences of a freely made choice, and (2) coding a prediction error using reward and prediction.

In an fMRI experiment on humans, Walton et al. (2004) pinpointed that ACC might use both rewards and errors to construct a history of choice-action associations. These associations can then be used to determine which future decisions should be taken (Walton et al., 2004). This role of ACC is supported by a single-unit recordings study by Kennerley et al. (2006). The authors could demonstrate that monkeys with ACC lesions were not impaired during reinforcement learning task with switched contingencies. However, the monkeys could not integrate value over time to adapt responses in a changing environment (Kennerley et al., 2006).

With regard to coding prediction error, it might seem that ACC neurons have an activity pattern similar to dopaminergic neurons. However, there are important differences. On the one hand, the same dopamine cell encodes both positive and negative prediction errors. On the other hand, as reported by Matsumoto et al. (2007) and Hayden et al. (2009) cells inside ACC have specific roles. Some ACC cells encode positive feedback, others encode their negative feedback or rewards that were not received (Matsumoto et al., 2007; Hayden and Platt, 2010).

Similarly, other single-unit recordings studies confirm that ACC is crucial for adaptation of action value (Quilodran et al., 2008), and linking behavioral regulations to monitor action valuation (Amiez et al., 2006). Lastly, this role of ACC is also reported by fMRI studies in healthy humans. Using a gambling task with changing reward state (i.e. high volatility), another study demonstrated that the BOLD signal in the ACC is correlated with the volatility estimate (Behrens et al., 2007). A more recent study by (Klein-Flugge and Kennerley, 2016) revealed that the network of regions involving dorsal ACC encodes the difference between the chosen and unchosen choices' subjective values.



© 2011 Pearson Education, Inc.

**Figure 1.16. Basal ganglia**

*A.* A frontal section of the brain showing the locations of the basal ganglia.

*B.* Lateral view showing the major nuclei of the basal ganglia.

Figure by RobinH at English Wikibooks - Transferred from en.wikibooks to Commons., CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=36970335>

### 1.7.2 Basal Ganglia

Basal ganglia (Figure 1.16), situated at the base of the forebrain, consists of several nuclei that includes the striatum, globus pallidus, ventral pallidum, substantia nigra, and subthalamic nucleus. The striatum is further divided into the dorsal striatum (caudate nucleus and putamen) as well as the ventral striatum (nucleus accumbens and olfactory tubercle). Due to its location, this subcortical structure is an excellent candidate for interactions with the executive functions. In particular, the striatum receives massive inputs from the neocortex.

The first evidence of the basal ganglia's role in reward and decision making comes from lesion studies in animal. It was shown that lesions of basal ganglia did not significantly affect fine movements. However, animals with basal ganglia lesions showed a decrease in actions needed to get specific reward (Konorski, 1967). Since then, many studies proved that neurons in several brain structures are sensitive to rewards. This is especially true for dopaminergic neurons. These neurons project from the substantia nigra, the ventral tegmental area, and the hypothalamus. They show short, phasic activation after presentations of rewards or after stimuli that predicts reward as well as an increase in activity for unpredicted rewards, i.e. reward prediction errors (RPE) (Schultz, 1998). This RPE signal is the key internal variable of model-free choice system in the sense of representing the avoidance of task structure. Thus, these studies became early precursors to the investigation of the model-free choice mechanism within the basal ganglia circuit.

In humans, one of the earliest investigation of the BOLD signal correlates of RPE in the ventral striatum was the study by O'Doherty et al. (2003). In particular, they observed a signed prediction error. In other words, a negative prediction error led to a negative BOLD signal. Further study by Seymour et al. (2004) demonstrated separate functions of the ventral and dorsal striatum. They found that the ventral striatum was mainly responsible for the structure learning to predict future outcome. In contrast, the dorsal striatum was specifically involved in maintaining information about the rewarding action outcomes (Seymour et al., 2004). This finding has been widely supported by a number of studies demonstrating correlates of RPE in the ventral striatum (Knutson and Cooper, 2005; Preusschoff et al., 2006; Hare et al., 2008; Preusschoff et al., 2008).

In 2010, however, Daw et al. (2011) showed that BOLD signal in the striatum reflected both model-free and model-based choices. Furthermore, the contribution of the model-free and the model-based valuations was found to be a proportional match to those that determine choice behavior. The authors concluded that rather than a separate brain regions for either choice system, the dopaminergic neurons support a more integrated computational architecture for value-based decision making (Daw et al., 2011).

Building on this idea, Wunderlich et al. (2012) designed a novel value-based decision task. This task consisted of two components, one for each choice system. The authors found that BOLD signals in the caudate pertained to individual computational components of model-based values. In contrast, signals in the posterior putamen selectively fluctuated with key internal variables of model-free choices. These results are direct evidence for multiple decision systems that operate independently and in parallel. These decision systems recruit neural structure along a mediolateral axis in the basal ganglia (Wunderlich et al., 2012).

Moreover, results from fMRI studies show that age-related changes in value-based decision making is prominently seen within the striatum (Chowdury et al., 2013; Eppinger et al. 2013). These two studies reported a reduced correlation between reward prediction errors (i.e. hallmarks of model-free choice) and BOLD signal in the ventral striatum. In line with this result, Samanez-Larkin et al. (2010) found that suboptimal financial decision making in older adults is correlated with increased temporal variability of the ventral striatal BOLD signal. Other studies point out that age-related changes in decision making may depend on the complexity of the decision environment. Older adults mostly rely on the use of decision strategy with shorter temporal horizon. They also performed worse than their younger counterparts whenever they had to rely on model-based choice (Mata et al., 2010; Worthy et al., 2011). Taken together, these results are consistent with several theoretical proposals suggesting age-related changes in value-based decision making might result from reduced dopaminergic projections from the mid-brain to the ventral striatum and vmPFC (Hämmerer and Eppinger, 2012).



## 1.8 Aim of the Thesis

As shown throughout this chapter, spatial navigation and wayfinding have been extensively studied from a variety of perspective including behavioral, electrophysiological, and brain imaging. These research efforts result in a wide range of findings: from the early breakthrough, such as those of Tolman (1948) that pointed toward internal representation of space, to the discovery of place cells (O'Keefe and Dostrovsky, 1971), head direction cells (Taube, Muller, and Ranck, 1990), and grid cells (Hafting et al., 2005). In human, various methods have been employed to discover the exact relation of brain structures with different elements of spatial navigation including strategy adoption. These studies provide insights into how the brain keeps track of our position as we move through space, and which brain region supports certain strategy adoption. Furthermore, each of these studies adds a piece of the multidimensional puzzle to solving navigational problems. In other words, each study contributes to the understanding on how the brain encodes, updates, manipulates, and uses spatial information.

This thesis aims to add an important piece to this puzzle by bridging two rather separate fields: spatial navigation and value-based decision making. By adding computational models based on RL algorithms into the neuroimaging analysis, neuroimaging data show not only which brain activity is correlated with either route-based or map-based navigation but also how that brain area may carry out navigation task on a trial-by-trial basis. The key internal variables, such as values of the RL algorithms, calculated at each time step were used to test different hypotheses about the possible mechanisms that the brain implement to drive strategy adoption. Thus, a substantial part of this work focuses on how computational modeling based on RL algorithms can help us better understand how the brain uses a cognitive map, assigns values to certain part of the environment, and uses this information to make decisions and navigate accurately. Attempting to understand these processes even further, this thesis also investigates how certain factors, namely age, affects individual and group differences in spatial navigation. The following chapters present two studies that were performed within the scope of this doctoral thesis.







# Neural Signatures of Reinforcement Learning Correlate with Strategy Adoption during Spatial Navigation

**Dian Anggraini**<sup>1,3\*</sup>, Stefan Glasauer<sup>2,3</sup>, Klaus Wunderlich<sup>1,3</sup>

1: Department of Psychology, Ludwig-Maximilians University of Munich

2: Center for Sensorimotor Research, Klinikum der Uni Muenchen

3: Graduate School of Systemic Neurosciences, Ludwig-Maximilians University of Munich

## **Authors' Contribution:**

Conceptualization: D.A., S.G., and K.W.; Methodology: D.A, S.G., and K.W.; Software: D.A. and K.W.; Validation: D.A, S.G., and K.W.; Resources: S.G.; Investigation: D.A.; Formal Analysis: D.A. and K.W.; Writing – Original Draft: D.A. and K.W.; Writing – Review and Editing: D.A., S.G., and K.W.; Visualization: D.A. and K.W.; Supervision: S.G. and K.W; Project Administration: D.A.; Funding Acquisition: S.G. and K.W.

\*D.A is the sole first author. The final version of this manuscript is currently under review and revision at journal Scientific Reports.

## 2.1 Abstract

Human navigation is generally believed to rely on two types of strategy adoption, route-based and map-based strategies. Both types of navigation require making spatial decisions along the traversed way. Nevertheless, formal computational and neural links between navigational strategies and mechanisms of value based decision making have so far been underexplored in humans. Here, we employed functional magnetic resonance imaging (fMRI) while subjects located different target objects in a virtual environment. We then modelled their paths using reinforcement learning (RL) algorithms, which successfully explain decision behaviour and its neural correlates. Our results show that subjects used a mixture of route and map-based navigation, and their paths could be well explained by the model-free and model-based RL algorithms. Furthermore, the value signals of model-free choices during route-based navigation modulated the BOLD signals in the ventro-medial prefrontal cortex (vmPFC). On the contrary, the BOLD signals in parahippocampal and medial temporal lobe (MTL) regions pertained to model-based value signals during map-based navigation. Our findings suggest that the brain might share computational mechanisms and neural substrates for navigation and value-based decisions, such that model-free choice guides route-based navigation and model-based choice directs map-based navigation. These findings open new avenues for computational modelling of wayfinding by directing attention to value-based decision, differing from common direction and distances approaches.

## 2.2 Introduction

When we navigate in daily life, we commonly do so with a goal in mind, such as going to a restaurant. To reach this destination, we can follow a route that we know from previous journeys. Alternatively, we can work out the shortest path using our cognitive representation of the neighbourhood. These two possibilities exemplify the route-based and map-based navigation respectively, two prominent strategy adoptions in wayfinding and spatial navigation (Iaria et al., 2003; Igloi et al., 2009; Wolbers, 2010). The route-based navigation, which is a form of response learning, relies on associations between landmarks and turns, as well as memory of travelled distances (Iaria et al., 2003; Bohbot et al., 2004; Latini-Corazzini et al., 2010 ). The medial prefrontal cortex, striatum, retrosplenial, and medial temporal regions appear to be important for route-based navigation (Aguirre and D'Esposito, 1997, 1999; Hartley et al., 2003; Wolbers et al., 2004; Latini-Corazzini et al., 2010). In contrast, a map-based or place navigation requires knowledge of the spatial relationship between goals, landmarks, or other salient points in space (Maguire et al., 1998; Shelton and Gabrieli, 2002; Wiener et al., 2013). This knowledge can be conceptualized as a cognitive map, which is defined as a mental representation of spatial environment. This representation enables one to acquire, store, code, and recall the relative locations as well as attributes of prior experience in that environment (Tolman, 1948; Redish, 1999; Eichenbaum and Cohen, 2014). The cognitive map is acquired by either active searching and exploration, or experiencing the environment using controlled navigational practices. Examples of controlled navigation practices include exploration using path integration and sequenced neighbourhood search (Golledge, 1999). Novel paths may be planned by first searching this map for the best path, and then translating this knowledge into a sequence of movements. Map-based navigation engages a distributed system of brain areas including hippocampus (Wolbers and Büchel, 2005; Iglói et al., 2010; Marchette et al., 2011; Brodt et al., 2016), parahippocampal (Epstein et al., 2007; Zhang et al., 2012; Urgolites et al., 2016), and retrosplenial regions (Miller et al., 2014b). These studies have identified which neural pathways are involved in each strategy. However, the computational mechanisms accounting for interaction of route and map-based navigation as well as its neural underpinnings remain a topic of debate, particularly in humans.

In this study, we address these questions by combining a 3D virtual environment, as typically used in imaging studies of spatial navigation, and modelling techniques from the field of value-based decision making. In contrast to previous studies that mostly answered where in the brain different spatial navigation strategies take place, our approach allows us to also investigate how these processes might be solved computationally. In addition, this approach also offers an experimental investigation into the utility of the modelling techniques from value-based decision making research in characterizing strategy adoption in spatial navigation.

Similar to the route-based and map-based strategies in spatial navigation, at least two complimentary systems have emerged as dominant behaviours for value-based decision making: model-free and model-based choice systems (Doya, 1999; Daw et al., 2005; Redgrave et al., 2010; Daw et al., 2011). In the model-free choice, associations are formed between stimuli and actions by reinforcing previously successful actions. Knowledge of the task structure is not necessary for the model-free choice behaviour. The model-based choice, on the other hand, relies on a cognitive model of the task structure to evaluate which sets of actions lead to the best outcome. This is done by searching through a map or graph of the task. Extensive neuroimaging studies have suggested computational principles for these two decision systems. These studies also elucidated putative corticostriatal circuits that underlie these computations (Glascher et al., 2010; Wunderlich et al., 2011; Hunt et al., 2012; Meder et al., 2017).

Since the computational principles underlying these two value-based choice systems successfully explain a multitude of decision behaviors, they may also provide a useful framework for spatial navigation. While links between spatial navigation and decision making have been previously suggested, mostly in investigations of animal wayfinding (van der Meer and Redish, 2011; Kurth-Nelson et al., 2012; Regier et al., 2015), these studies addresses more general questions than the one addressed by our hypothesis. We propose that the brain shares some of the key computational mechanisms for navigation and for making value-based decisions in a way that model-free choice guides route-based navigation while model-based choice directs map-based navigation. Therefore, we used reinforcement learning (RL) algorithms to model subjects' traversed paths in the virtual environment. If RL algorithms form the basis of navigational strategies, they should (1) account for subjects' behaviors when they are free to choose their own navigational



strategy, (2) distinguish the degree to which one system is used at each decision point, and (3) explain the role of different brain regions in processing navigational decisions by showing a correlation between BOLD activity and key internal variables of the RL algorithms.

To test this hypothesis, we created a wayfinding task in a 3D virtual environment (VE) for human subjects. Functional magnetic resonance imaging (fMRI) was then used to investigate how blood oxygen level dependent (BOLD) signal is modulated by computational processes while subjects performed the wayfinding task. The 3D environment consisted of a 5 by 5 grid of rooms. Each room was distinctively furnished to allow subjects to distinguish individual rooms (see Figure 2.1. A and B). Every subject performed three phases of wayfinding tasks in the same environment. Subjects navigated by freely choosing one of the available doors. Since backtracking (leaving the room from the same door as entering it) was not allowed, rooms in the middle of the grid-world had three doors to choose from, rooms along the outside wall had two doors, and corner rooms had one. First, in the (1) encoding phase, subjects collected three rewards in the same predefined order and from the same starting position over eight trials. This repetition encouraged route-learning during the encoding phase. During the subsequent (2) retrieval phase subjects collected one randomly chosen reward in each trial from the same initial position as in the encoding phase. Lastly, during the (3) search phase, subjects collected one randomly chosen reward at a time, from a different starting location. These phases, especially the retrieval and search phases, allowed subjects to use their preferred navigational strategy (further details on the wayfinding task are provided in the Supplementary Method). By analyzing the paths that the subjects took, we calculated three navigation indices. These indices describe the degree to which each subject followed previously learned routes (route-based navigation) or engaged in map-based navigation. Based on reinforcement learning models, we also fitted a parameter ( $\omega$ ) that explains whether subjects relied more on model-free or model-based choices during navigation.

In our task, our expectation was that subjects would use a mixture of route and map-based navigation across trials, and that their traversed paths could be well explained by the model-free and model-based RL algorithms. Significant correlations between navigation indices and  $\omega$  would then confirm that RL algorithms account for individuals'

variability in strategy adoption during navigation. Moreover, we expected that correlates between BOLD activity and RL's key internal variables, such as value signals, from the model-free and the model-based choice systems may provide evidence that these algorithms explain how the brain computationally solves the navigation problem.

## 2.3 Results

### 2.3.1 Strategy Adoption during Wayfinding Task

During the early trials of the encoding phase, most participants started by exploring the environment. This was often seen by moving from one side of the maze to the opposite until all rewards were collected. Then subjects either found the shortest paths to go between one reward and the next or established a certain route to go from one reward to the other (Figure 2.1. C). During retrieval and search phases, some subjects (16 out of 27) used the optimal path at least 60% of the time to retrieve rewards. Others chose to follow the route they established during the earlier encoding phase. The latter was a clear sign of employing route-based navigation. Interestingly, during the search phase, when some subjects started on part of an established route they would simply follow it to reach the reward. However, when they started from a position that was not part of their established route, they located the reward using the shortest path. Those subjects might have constructed a cognitive map of the environment but used it only when required to plan a new path. Otherwise, they relented to the cognitively less demanding but potentially longer route strategy. Note that no subject chose exclusively one strategy over the other.

As a first crude measure to quantify strategy adoption in our subjects, we calculated three navigation indices based on: (1) the number of trials in which the shortest path was used (adapted from (Marchette et al., 2011) and referred to as  $I_{PATH}$  in this work), (2) the excess number of steps in comparison to the shortest path required to reach the target (referred as  $I_{STEPS}$ ), (3) the number of repeated route trials (referred as  $I_{ROUTE}$ ). Each navigation index was first calculated individually for each phase of the wayfinding task. For the search phase, we could distinguish whether subjects used novel shortest paths to retrieve rewards or simply followed the route they established during the earlier encoding phase. This distinction was sometimes not so clear cut for the retrieval phase because subjects had to find each reward from the same starting position across several trials. Nevertheless, across three different phases, we still had enough trials to calculate indices that quantified strategy adoption.

Therefore, to get an overview on how subjects performed throughout the entire experiment, every navigation index was then averaged across three phases of the wayfinding task. For both  $I_{PATH}$  and  $I_{STEPS}$ , a value of 1 indicates that a subject would have used the shortest path on every trial. In other words, the subject was primarily displaying a map-based strategy. As the definition suggested, we expected a significant correlation between  $I_{PATH}$  and  $I_{STEPS}$  ( $R^2 = 0.85, P < 0.01$ ). This is because, by design, only subjects who had a good spatial representation of the environment could have reached the rewards using the shortest paths. In contrast, subjects whose scores were close to 0 predominantly used suboptimal long routes, suggesting that they lacked a map-like representation of the environment.

Since  $I_{PATH}$  and  $I_{STEPS}$  were calculated based on the trials when subject used optimal paths to locate rewards, one might argue that a low score (closer to 0) might reflect a lower performance level rather than an indication of subjects using route-based strategy. To address this issue and ensure that our measured are not confound by overall navigational success, we also calculated the index  $I_{ROUTE}$ . This index was calculated as a proportion of trials where subjects simply repeated previously learned routes and trials with the shortest path. A score of 1 in this index suggest that a subject kept repeating the same route throughout the experiment. A significant correlation between  $I_{PATH}$  and  $I_{ROUTE}$  ( $R^2 = 0.88; P < 0.01$ ) as well as  $I_{STEPS}$  and  $I_{ROUTE}$  ( $R^2 = 0.79, P < 0.01$ ) suggested that a lower score on both  $I_{PATH}$  and  $I_{STEPS}$  did not come exclusively from low performance.

The distribution of different values of  $I_{PATH}$ ,  $I_{STEPS}$ , and  $I_{ROUTE}$  revealed that although there are some subjects that showed strong tendency toward either route or map-based navigation, the majority fell in between (Supplementary Table 2.1). Further details on the indices are provided in the Methods and the further details of navigation indices in each phase of the wayfinding task are provided in the Supplementary Information.

To measure subjects' general spatial cognitive ability, subjects performed a paper-based Mental Rotation Test (MRT). We found significant correlations between the score on the MRT ( $M = 25.63, SD = 9.43, N = 25$ ) and  $I_{PATH}$  ( $M = 0.42, SD = 0.19, r(25) = 0.72$ , two-tailed t-test  $P < 2.11 \times 10^{-5}$ ),  $I_{STEPS}$  ( $M = 0.68, SD = 0.22, r(25) = 0.72$ , two-tailed t-test  $P < 1.95 \times 10^{-5}$ ) as well as  $I_{ROUTE}$  ( $M = 0.47, SD = 0.18, r(25) = -0.72$ , two-tailed t-test  $P < 2.44 \times 10^{-5}$ ). These results indicate that greater ability to mentally represent and

manipulate objects was associated with higher capability in planning shorter paths. These results are somewhat expected as object-based spatial ability (e.g. mental rotation, spatial visualization) is one of the prerequisites for good performance in map-based navigation (Hegarty and Waller, 2005).

### 2.3.2 RL Model Fits to Navigation Data

Similar to navigation in real-life, our wayfinding task is directed toward certain goals and requires decisions about directions along the traversed path. Subjects who used the shortest path could find the reward rooms more quickly and thereby accumulated more rewards. To computationally assess subjects' navigation strategy at every decision point, we modelled subjects' choice behaviour by fitting three different reinforcement learning (RL) algorithms: model-free, model-based, and a hybrid model. The hybrid model is formulated as a weighted combination of model-free and model-based algorithms. These algorithms exemplify two strategies in value-based decision making: (1) the model-based choice that creates a cognitive representation of the entire environment, and (2) the model-free choice that simply increases action values along the taken paths that previously led to rewards (Figure 2.2). The hybrid model assumes that subjects would employ both algorithms at a relative degree, represented by a fitted parameter weight ( $\omega$ ) (Glascher et al., 2010; Daw et al., 2011). We performed model fitting for each subject individually and assessed the relative goodness of fit in every phase (for details see Methods, Supplementary Methods, and Supplementary Table 2.2).

The fitted  $\omega$  (first fitted for every phase then averaged over three different phases) in the hybrid model ( $M = 0.50$ ,  $SD = 0.23$ ,  $N = 27$ ) significantly correlated with spatial cognitive ability (as assessed by MRT,  $r(25) = 0.54$ , two-tailed t-test  $P < 0.0035$ ) both  $I_{PATH}$  ( $M = 0.42$ ,  $SD = 0.19$ ,  $r(25) = 0.88$ , two-tailed t-test  $P < 8.28 \times 10^{-10}$ ) and  $I_{STEPS}$  ( $M = 0.68$ ,  $SD = 0.22$ ,  $r(25) = 0.80$ , two-tailed t-test  $P < 4.69 \times 10^{-7}$ ). These results demonstrate that, across three phases of the wayfinding task, subjects who often took the shortest paths or a relatively small number of steps also showed a tendency towards the model-based choice (Figure 2.3. A and B). In line with these results, we also found significant negative correlation between  $\omega$  and  $I_{ROUTE}$  ( $M = 0.48$ ,  $SD = 0.18$ ,  $r(25) = -$

0.82, two-tailed t-test  $P < 1.52 \times 10^{-7}$ ). This shows that subjects who repeated learned routes are better explained by model-free RL (Figure 2.3. C).

### 2.3.3 Neural Signatures of Model-Free and Model-Based Choices

We then investigated neural responses pertaining to choice valuations in every room. A key internal variable of the RL algorithms is the value of the chosen action, a signal that has been reliably detected in BOLD fluctuations over a large number of studies (Wunderlich et al., 2009; Glascher et al., 2012; Kahnt et al., 2014). For every decision point, we took the values of the chosen action (i.e. go to the left, right, or straight ahead), calculated separately based on model-free and model-based RL, along the traversed path as parametric modulators for our fMRI data. By analysing 3 different phases of the wayfinding task in one GLM, we found that BOLD activity correlated significantly with the model-free value signals along left ventromedial prefrontal cortex [vmPFC, peak:  $x = -3, y = 47, z = -13$ ] extending to left anterior cingulate cortex [ACC, peak:  $x = -3, y = 38, z = 5$ ]. Other clusters included retrosplenial complex [ $x = 6, y = -52, z = 32$ ] and caudate nucleus [ $x = -9, y = 11, z = 14$ ]. In contrast, significant correlations with the time series of model-based values were most prominently in the area of right parahippocampal gyrus [peak:  $x = 21, y = -46, z = 2$ ] extending to hippocampus and bilateral calcarine gyrus. Additionally, activity in right precuneus [ $x = 27, y = -64, z = 29$ ] and left retrosplenial complex [ $x = -12, y = -31, z = 44$ ] correlated with model-based values. See Figure 2.4 for the activated areas and Supplementary Table 2.3 and Supplementary Table 2.4 for a list of all activated areas. For correction of multiple comparisons, we set our significance threshold at  $P < 0.05$  whole-brain FWE corrected for multiple comparison at cluster level.

These results show that our behavioural models explain subjects' path choices and subjects' brain activity represents crucial decision variables in this process. Therefore, we would expect that brain activity should be particularly well-explained in those subjects in whom our model also provides a good choice prediction. Specifically, model-free values should explain BOLD activity in vmPFC particularly well in those subjects for whom our hybrid model indicated a large route-based contribution. Similarly, brain activity in the parahippocampal/hippocampal area should be particularly well explained

by the model-based values in those subjects in which the hybrid model indicated a large map-based contribution. Across subjects this would be expressed in a relationship between strategy choice (represented by the  $\omega$  parameter) and strength of the activity pertaining to the parametric value signals of the corresponding model ( $\beta$ -parameter estimates in the general linear model) of the fMRI analysis. Consistent with our conjecture, we found a significant negative correlation ( $r(25) = -0.386$ , two-tailed t-test  $P = 0.046$ ) between  $\omega$  from the hybrid model and  $\beta$  estimates in ventromedial PFC for model-free value signals (Figure 2.5. A). Note that a smaller  $\omega$  indicates a larger degree of route-based influences. As expected, we did not find a correlation ( $r(25) = -0.081$ , two-tailed t-test  $P = 0.687$ ) between  $\omega$  and the  $\beta$  estimates for the model-based value signals in the medial PFC (Figure 2.5. A). Similarly, we found a significant positive correlation ( $r(25) = 0.404$ , two-tail t-test  $P = 0.036$ ) between  $\omega$  and the extracted parameter estimates of the model-based value signals in right parahippocampal gyrus (Figure 2.5. B). Again, we did not find a significant correlation ( $r(25) = -0.0029$ , two-tailed t-test  $P = 0.988$ ) between  $\omega$  and parameter estimates of the model-free  $\beta$  estimates in the parahippocampal area (Figure 2.5. B).

Furthermore, these correlations confirms that our RL algorithms explain a larger proportion of the fluctuation in the neuronal data in those subjects whose choices are well explained by the respective model. That is, the more subjects lean towards one navigation strategy, the more clearly the computational value signals of the corresponding decision mechanisms are seen in either the vmPFC or the parahippocampal area. Note that we did not claim that there is a significant difference between the two correlations in both parahippocampal gyrus and mPFC. What we showed was that, across subjects and within a certain brain region, there is a selective relationship between strategy choice on the behavioural level and the representation of the corresponding value signal in the brain. In effect, these results show that BOLD activity in vmPFC and parahippocampal area reflected model-free or model-based valuations in proportion matching those that determine subjects' use of route-based or map-based navigation.

We also tested several alternative hypotheses for putative signals that the brain might keep track of during navigation. One such hypothesis is that the brain combines values

from the model-free and model-based using the relative weight parameter  $\omega$ . We tested for this by looking for neural signals pertaining to values of the hybrid RL model, which is a weighted combination of model-free and model-based values. These signals did not survive cluster level correction for multiple comparison.

We further explored the idea that the brain encodes the distance to goals during navigation. We found that while activity in the precuneus positively correlated with the distance to goal, activity in fusiform and caudate increased as subjects approach the goal. Consistent with previous studies, activity in precuneus may be indicative of a spatial updating process during navigation (Wolbers et al., 2008; Spiers and Barry, 2015). On the contrary, activity in fusiform gyrus and caudate nucleus increases as subjects approach the goal. Based on a previously demonstrated role of these regions in object recognition (Ekstrom and Bookheimer, 2007; Grill-Spector et al., 2006), it is conceivable that participants were more engaged in recognizing the object in the room to mentally simulate the goal's location as they were getting closer to the goal. Furthermore, while some studies suggest that BOLD signal in the entorhinal cortex (Spiers and Maguire, 2007) and hippocampus (Howard et al., 2014) pertained to goal distance tracking, we did not find any clusters within the hippocampus and entorhinal cortex that are correlated either positively or negatively with distance to goal. This absence of activity, similar to the one reported by Slone et al. (2016), may suggest that the type of environment (network of rooms, open vista, or city space) affects distance-tracking neural circuitry. It is important to note that our finding of neural correlates of model-based value signals in the medial temporal region was still present when we accounted for the distance to goal as a potential confound. Our results suggest that medial temporal regions not only encode proximity or distance to the goal but may also represent value calculations during map search and wayfinding.

Because there might be a functional relationship between distance to goal and model-based choice values, we also tested if our earlier results for model-based choice values are indeed correlates of value signals, or could potentially be only a spurious result of this relation. To test for this we estimated a GLM containing both model-based value and distance parametric regressors without orthogonalization, letting them directly compete for variance. Any remaining activation in this analysis can be attributed uniquely to one or the other regressor. Importantly, we still found a significantly correlated BOLD



activity with model-based value signals in medial temporal region in this design (see Supplementary Figure 2.4).

## 2.4 Discussion

We show that human subjects are adept at flexibly employing route-based and map-based strategies while navigating in a wayfinding task. Their choices in each room correlated well with the fitted parameter  $\omega$  from the hybrid model, which is a mixture of model-free and model-based RL algorithms. The computational processes of model-free and model-based RL were also represented at the neuronal level. Model-based valuation during map-based navigation strongly modulated activity in parahippocampal and medial temporal lobe (MTL) areas. In contrast, BOLD responses in striatum and vmPFC pertained to model-free valuations. Equally important, we found a direct link between the degree to which subjects used one navigation strategy and the neural representation of value signals associated with the corresponding decision mechanisms.

Converging evidence from both animal and human studies has suggested that navigation predominantly relies on two types of strategy adoption generally termed as route-based and map-based navigation (Ekstrom et al., 2014; Wolbers and Wiener, 2014). In most previous studies on spatial navigation subjects were constrained to one navigation strategy, such as passively following a certain route during the encoding phase (Wolbers et al., 2004; Latini-Corazzini et al., 2010; Wiener et al., 2013). Our encoding phase was different in that we allowed subjects to explore. This feature is helpful to elucidate the computational mechanisms accounting for route and map-based navigation. Because subjects were not constrained to one strategy over the other, we could observe the use of both strategies in the same environment. Even though our encoding phase emphasized the formation of a route between a single starting position and a sequence of three reward locations, it simultaneously allowed the formation of a map-like representation over time because subjects could freely move within the environment. The retrieval phase then allowed further memory formation of how every target location could be reached from the starting position.

To our knowledge, one previous neuroimaging study on value-based decision making has attempted to characterize neural and computational substrates of reinforcement learning by incorporating some spatial elements (Simon and Daw, 2011). Albeit the task created for their study adopted a grid-world configuration commonly used in human navigation, their research questions were more concerned with value-based decision

making. For instance, their task included visual cues that constantly informed subjects about the location of the target. Moreover, throughout the experiment, there was an ongoing reconfiguration of the grid-world and occasional random teleportations of the subject. These features are useful to study neural mechanisms of reinforcement learning but preclude investigation of wayfinding using optimal or repeated routes. Consequently, in comparison to this study, our task mimics more of an everyday experience. This allows us to study human spatial navigation where subjects can combine and switch between different strategies depending on which is optimal in each situation.

The properties and neural correlates of route and map-based strategies have been intensively investigated in both humans (Maguire et al., 1999; Ekstrom et al., 2014) and animals (Moser et al., 2008). Although the terminology has evolved, the route-map dichotomy is still valid. Previous studies have also investigated the emergence of route vs. map knowledge and suggested that landmark knowledge (necessary for route navigation) and survey knowledge can be acquired in parallel (Newman et al., 2007). In addition, preferential engagement of route and map-based navigation may be a dimension along which individuals vary (Marchette et al., 2011; Marchette et al., 2014). In the neuroimaging literature, some results strongly support the route-map dichotomy (Hartley et al., 2003; Iaria et al., 2003; Wolbers et al., 2004; Wolbers and Büchel, 2005; Iaria et al., 2007; Wolbers, 2010). Other data reveals overlapping or common activations, suggesting that in some cases, the dichotomy may not be so clear cut (Shelton and Gabrieli, 2002; Nori and Giusberti, 2006; Igloi et al., 2009; Latini-Corazzini et al., 2010). Our data supports not only a functional separation in wayfinding between medial prefrontal cortex for route based and medial temporal areas for map based strategies but also include common activations such as in the retrosplenial complex.

In the parahippocampal and MTL area, we observed activations related to choice values that are likely correlates of model-based choice computations. The hippocampus has been shown to play a key role in processing relative spatial and contextual information, as well as in encoding cognitive maps of the spatial environment and the current position in space in both animals (O'Keefe and Nadel, 1978; Redish, 1999; Pfeiffer and Foster, 2013) and humans (Ekstrom et al., 2003; Doeller et al., 2008; Jacobs et al., 2013). In our wayfinding task, these processes are crucial for forming a map-like representation of the environment. To do this, subjects need to not only keep track of their current location

relative to the reward location but also to integrate and transform spatial information into survey knowledge (or an allocentric reference system). Moreover, the parahippocampal gyrus responds selectively to visual scenes depicting places (Epstein and Kanwisher, 1998) and is also specifically involved in the retrieval of spatial context compared with non-spatial context (King et al., 2002; Bar et al., 2008; Wolbers et al., 2011). In addition, this brain areas is also crucial for identification and retrieval of landmarks (O'Craven and Kanwisher, 2000; Burgess et al., 2002), as well as spatial relationship and relevance of landmarks encountered during navigation (Hartley et al., 2003; Janzen and van Turenhout, 2004; Rosenbaum et al., 2004; Rosenbaum et al., 2015). A coherent map-like representation of a complex environment, such as the one used in the present study, requires the ability to identify and retrieve different landmarks, as well as to form links between different landmark identities and the layout of local areas. Consequently, model-based computations in the parahippocampal and MTL area during our wayfinding task could be indicative of searching and planning within a mental representation of the virtual environment to find the shortest path to reach the goal.

For the neural correlates of model-free choices, we observed activations in vmPFC and caudate nucleus, which are consistent with previous studies in animals and humans (Daw et al., 2005; Doeller et al., 2008; Chersi and Burgess, 2015). Our model-free RL algorithm assigns values to subjects' choices along the taken path. This approach emphasizes the character of route-based navigation that includes temporal relations between landmarks and sequences of turns (Hartley et al., 2003; Wolbers et al., 2004; Foo et al., 2005; Igloi et al., 2009). In addition, different studies have proposed that vmPFC encodes stimulus-reward associations (Hampton et al., 2006; Rushworth and Behrens, 2008; Balleine and O'Doherty, 2010; Basten et al., 2010; Daw et al., 2011; Apps and Ramnani, 2014). Thus, by representing model-free valuation, the brain may recall previously cached landmark-action associations. Equally important, activity in the vmPFC has also been observed in spatial working memory tasks as well as during retrieval of information about order and context (Kolb et al., 1994; Kaplan et al., 2014; Shikauchi and Ishii, 2015; Moscovitch et al., 2016). In rats, the dorsolateral striatum (DLS) has been shown to be involved in a spatial learning process that relies on rewarded stimulus-response behavior, i.e. model-free choices (van der Meer and Redish, 2011; Ferbinteanu, 2016). In other words, neural activity in this region correlates with learning of turns and response to stimuli, which is the hallmark of route-based navigation. In humans, several fMRI studies report that the

caudate is activated during route-following and route-recognition tasks (Iaria et al., 2003; Wolbers et al., 2004; Etchamendy and Bohbot, 2007; Latini-Corazzini et al., 2010; Marchette et al., 2011). Thus, we argue that the encoding of model-free valuations in the vmPFC and left caudate nucleus during our navigation study may reflect computational mechanisms necessary for encoding of relations between landmarks as well as tracing of sequence of turns and places, which are crucial for route-based navigation.

The retrosplenial involvement in navigation has been demonstrated in electrophysiological, neuroimaging, and lesion studies (Ino et al., 2002; Wolbers et al., 2004; Wolbers and Büchel, 2005; Epstein et al., 2007). This area is anatomically closely linked to various medial temporal regions and mid-dorsolateral prefrontal cortex. Patients with retrosplenial lesions have been reported to be unable to form or recall links between landmark identity and couldn't derive navigational information from landmarks (Miller et al., 2014a). These two processes are crucial to route-based navigation. Moreover, neuroimaging studies reported performance dependent activation in retrosplenial complex (RSC) during mental navigation (Ino et al., 2002). Other studies suggested the correlation of retrosplenial activation with the amount of survey knowledge acquired following learning the spatial relation in an environment (Wolbers et al., 2004; Wolbers and Büchel, 2005). These findings, along with the modulated activity in this region by both model-based and model-free regressors, support its prominent role in processing landmark information and using landmarks to navigate and discern space. This role is important for both route and map-based navigation.

Since our subjects were not confined to one navigation strategy or the other, they could apply both strategies during the experiment. One might speculate that using such a mixture could result in integrating information from both model-free and model-based computations in a weighted manner for every choice. In our experiment, however, we did not find direct evidence for such an integration as there was no significant correlation between value signals of the hybrid model and BOLD activity. This would support the hypothesis that route and map-based evaluations are not necessarily integrated on every decision point. Instead, we suggest there were trials when subjects relied heavily on either route or map-based strategy to guide their choices during navigation. Note, however, that by design, our hybrid model cannot finally discriminate between both hypotheses. Either way, our calculation of a behavioural weight of model-based versus model-free

computations is still meaningful since  $\omega$  is fitted over all trials in a given phase and indicates an aggregated relative degree of model-based versus model-free choices over the course of the experiment.

In our experiment, the target rooms that subjects had to find contained a reward (subjects were paid according to how many rewards they found) that we could use to update RL values. Therefore, one could argue that we only observed RL learning because of the rewarding nature of the task and this process might not be causally linked to making navigational choices. To rule out the influence of an external monetary reward on our results, we conducted an additional behavioural experiment. The wayfinding task was similar to the one presented here but without monetary rewards. We found no differences in subjects' choice or navigation behaviour (see Supplementary Results, Supplementary Methods, and Supplementary Figure 2.3 for details and results). On a neural level, even in situations without explicit reward, goal-directed navigation requires the brain to represent the goal location as some form of intrinsic reward, and the brain employs the same reward learning machinery whether rewards are real or fictitious (Lin et al., 2012; Boedecker et al., 2013). Moreover, if our results were only due to the value of reward, we would expect that the observed brain activations would only mirror the one found in pure decision making. Here we find BOLD activity pertaining to value-based computations in the medial temporal and retrosplenial regions. These regions are not commonly associated with RL but frequently reported during spatial navigation. Taken together, it is very likely that our findings indeed form the bases of wayfinding towards the goal state.

In conclusion, our findings show that successful navigation requires the ability to flexibly integrate different strategies depending on the given situation. Importantly, our results also suggest that the neural computations during employment of these strategies might be algorithmically described by RL models. Such a link between value-based decision making and spatial behaviour seems plausible given the central role that choice has played in navigation from the beginning of the evolution of the central nervous systems.

## 2.5 Methods

### 2.5.1 Participants

27 right-handed females, 20 – 29 years of age, participated in the wayfinding task while undergoing functional Magnetic Resonance Imaging (fMRI). Prior to the fMRI experiment, we conducted an extensive pilot study where we tested both female and male participants. The results of our pilot study showed that more than half of the male participants reached performance ceiling effects at the given level of task complexity. Instead of varying task complexity (e.g. size of the labyrinth), we chose to recruit only female participants to reduce variance in performance and strategies across subjects and thereby increase power of the tests. This is a common practice in psychological experiments that are designed to study general population mechanism compared to individual differences within the population (Hanel and Vione, 2016). While this may limit generalizability of our findings here to men, we are currently conducting a follow up study that includes participants across different gender and age groups that will address these issues.

Two additional participants were not included in the analysis because they did not complete all trials of the experiment. All participants had normal or corrected-to-normal vision and no history of either neurological/psychiatric illness or any other contraindications to the MRI environment. The study was approved by the ethics committee of the Medical Faculty of the Ludwig-Maximilian-University Munich. All participants gave written informed consent and were paid a compensation of 25 to 30 EUR based on the number of collected rewards during the experiment.

### 2.5.2 Navigation Indices

To quantify strategy adoption during the wayfinding task, we calculated three navigation indices. For each navigation index, we calculated the index separately for each phase of the wayfinding task (adapted from (Marchette et al., 2011)). We first measured (1) the number of trials when subjects used the shortest path to find the rewards, (2) the steps taken to complete one trial in excess of the minimal number of steps, and (3) number of trials where subjects repeated learned routes.

***I<sub>PATH</sub>* and *I<sub>STEPS</sub>***

We calculated the proportion of trials with the shortest path to the total number of trials in a respective run

$$I_{PATH} = \frac{n_{shortest\_trial}}{n_{total\_trial}} \quad [2.1]$$

$n_{shortest\_trial}$  is the number of trials when subjects used the shortest path to find the rewards and  $n_{total\_trial}$  is the number of trials in a given phase.

Alternatively we determined a navigation index based on calculating the excess steps in each phase. We then defined this index as

$$I_{STEPS} = \frac{n_{max\_steps} - n_{obs\_steps}}{n_{max\_steps} - n_{shortest\_steps}} \quad [2.2]$$

$n_{max\_steps}$  is the maximum number of steps found in each phase across all subjects,  $n_{shortest\_steps}$  is the number of shortest steps to reach reward (i.e. number of steps in optimal path),  $n_{obs\_steps}$  is the number of steps a subjects actually took. A score of 1 on both indices suggests that a subject used the shortest path on every trial, i.e., the subject was primarily displaying a map-based strategy. In contrast, a score of 0 indicates that a subject always used a suboptimal longer path.

***I<sub>ROUTE</sub>***

In both  $I_{PATH}$  and  $I_{STEPS}$ , a score of 0 does not necessarily indicate that subjects simply followed previously established route to reach rewards. To address this issue, we also calculated the proportion of trials with the shortest paths and trials where subjects simply repeated previously learned routes. Note that for the encoding phase, this index might be less meaningful because subjects were still learning the environment and their task was to establish a certain route. We defined this index as

$$I_{ROUTE} = \frac{n_{route\_trial}}{n_{route\_trial} + n_{shortest\_trial}} \quad [2.3]$$



$n_{route\_trial}$  is the number of trials when subjects repeated learned routes.  $I_{PATH}$ ,  $I_{STEPS}$ , and  $I_{ROUTE}$  were first calculated for each phase separately. To get a general idea on how each participant navigate throughout the entire experiment, we averaged each index over three phases of the wayfinding task.

### 2.5.3 Reinforcement learning model

We modelled the sequence of subjects' choices ( $a_i$ ) by comparing them step by step to those predicted by different learning algorithms as having encountered the same state ( $s$ ), action ( $a$ ) and reward ( $r$ ). As we had 5 by 5 grid, the wayfinding task consisted of 25 states and in each state, subjects could have up to three actions depending on which direction the subject was facing.

Because the wayfinding task consisted of three rewards, the goal for both model-free and model-based algorithms is to learn the state-action value function  $Q(s,a)$  mapping at each state-action pair to each reward. We assume no interference or generalization between the three rewards conditions, and thus each algorithm was subdivided into three independent task set, one for each reward.

#### ***Model-free reinforcement learning***

For model-free choice, we used the SARSA with eligibility traces (SARSA ( $\lambda$ )) to calculate model-free value or  $Q_{MF}$  (Sutton and Barto, 1998). This algorithm has three free parameters: learning-rate ( $\alpha$ ), inverse temperature ( $\beta$ ) and eligibility parameter ( $\lambda$ ). Each state-action pair is associated with a value  $Q_{MF}(s,a)$  all initially set to 0. The eligibility trace  $Z$ , set to 1 at the beginning of the trial and assumed not to be carried over from trial to trial, allows us to update each state-action pair along a subject's encountered trajectory. For every trial  $t$  in which the subject located the reward ( $r$ ), the state-action value is updated for each step  $i$  in that trial according to the following:

$$Q_{MF}(s_{i,t+1}, a_{i,t+1}) \leftarrow Q_{MF}(s_{i,t}, a_{i,t}) + \alpha \delta_{i,t} Z_i,$$

where

$$\delta_{i,t} \leftarrow R + Q_{MF}(s_{i+1,t}, a_{i+1,t}) - Q_{MF}(s_{i,t}, a_{i,t}),$$

and

$$Z(s_{i,t}, a_{i,t}) \leftarrow \lambda Z(s_{i,t}, a_{i,t}).$$

**Model-based reinforcement learning**

A model-based choice learns the configuration of the grid world and computes action values by searching across possible trajectories to locate the reward (Simon and Daw, 2011). Based on the grid-world configuration, we compute state-action values based on a planning process terminating at reward states. Specifically, for each action  $a$  in room  $s$ , we first initialized all  $Q_{MB}(s, a)$  to 0. Then, for all state-action pairs  $(s, a)$  and adjacent (next room) state-action pairs  $(s', a')$  we iteratively perform the following:

$$Q_{MB}(s, a) \leftarrow \begin{cases} R(s') & \text{if } R(s') \neq 0 \\ \max_{a' \in A} Q_{MB}(s', a') - (\gamma * \max_{a' \in A} Q_{MB}(s', a')) & \text{otherwise} \end{cases} \quad [2.7]$$

The algorithm has one fixed parameter  $\gamma$  that is set to 0.1. We took model-based values ( $Q_{MB}$ ) to be the values resulting after the algorithm converged (this occurred within 25 iterations). Note that computations of model-based value did not depend on the trial  $t$  or step  $i$  of the subject.

**Hybrid model**

In addition to model-free and model-based algorithm, we also considered a hybrid model (Glascher et al., 2010; Wunderlich et al., 2012) in which the model predicted values for the actions are calculated as a weighted linear combination of the values from model-free and model-based algorithms:

$$Q_{hybrid} = (1 - \omega)Q_{MF} + \omega Q_{MB} \quad [2.8]$$

The relative degree that the model-based algorithm contributed over the model-free is captured by the weight parameter ( $\omega$ ). We took this  $\omega$  as a free parameter, which was fitted individually for each subject but assumed to be constant throughout a single phase of the wayfinding task.

**Model fitting**

For each algorithm, we estimated a set of free parameters separately for each subject and for each phase of the wayfinding task by mean of hierarchical model fitting (Wunderlich

et al., 2012). Further details on the model fitting and calculation of model evidence are provided in the Supplementary Methods.

#### **2.5.4 fMRI Data Analysis**

The following software were used for the fMRI data analysis:

SPM12 (<https://fil.ion.ucl.ac.uk>, RRID:SCR\_007037);

Xjview (<https://alivelearn.net/xjview/>, RRID:SCR\_008642);

SPM Anatomy Toolbox (<https://fz-juelich.de/>, RRID:SCR\_013273);

mricron (<https://nitrc.org/mricron>, RRID: SCR\_002403).

#### ***General Linear Model (GLM) for fMRI data analysis***

An event related analysis was applied on two levels using the general linear model approach as implemented in SPM12. Individual (random-effects) model parameters were used to generate regressors for the analysis of the fMRI data. The GLM included the following event (time) regressors covering the time when subjects saw (1) the instruction, (2) the room, (3) chose which direction they wanted to go (button press), (4) animation of movement, and (5) seeing the reward. Our analysis focused on the times when subject entered each room and the button press to indicate where to go next. For our primary hypothesis, the decision time points were parametrically modulated by (1) model-free values, and (2) model-based values. Parametric regressors were not serially orthogonalized, thus allowing each regressor to account independently for the response at each voxel. Using this approach, we let the model-free and model-based value regressor directly compete for variance in the BOLD signal. In this approach, only variance that is exclusively explained by one or the other regressor is assigned to the regressor but not the variance that is shared by both.

All regressors were convolved with the canonical hemodynamic response function as provided by SPM12 and its temporal derivative. The six rigid-body motion parameters from the head motion correction were also included in the model as regressors of no interest. At the first level, linear weighted contrasts were used to identify effects of interest, providing contrast images for group effects analysed at the second (random-effect) level.

*Second level analysis.* Calculated linear contrasts of parameter estimates, from the first level GLM analysis, for each regressor were then brought to the separate second level random-effects analysis. At this level, one sample t-test provided effect for each regressor of interest. For correction of multiple comparisons, we set our significance threshold at  $P < 0.05$  whole-brain FWE corrected for multiple comparison at cluster level. The minimum spatial extent,  $k = 25$ , for the threshold was estimated based on the underlying voxel-wise  $P$  value with a cluster defining threshold of  $P = 0.001$ .

Details on the fMRI preprocessing are provided in the Supplementary Methods.

### **2.5.5 Region of Interest Analysis**

We extracted data for all region of interest analyses using a cross-validation leave-one-out procedure: we re-estimated our second-level analysis 27 times, always leaving out one subject. Starting at the peak voxel for the value signals in mPFC and right parahippocampal gyrus we selected the nearest maximum in these cross-validation second-level analyses. Using that new peak voxel, we then extracted the data from the left-out subject and averaged across voxels within a 4 mm sphere around that peak.

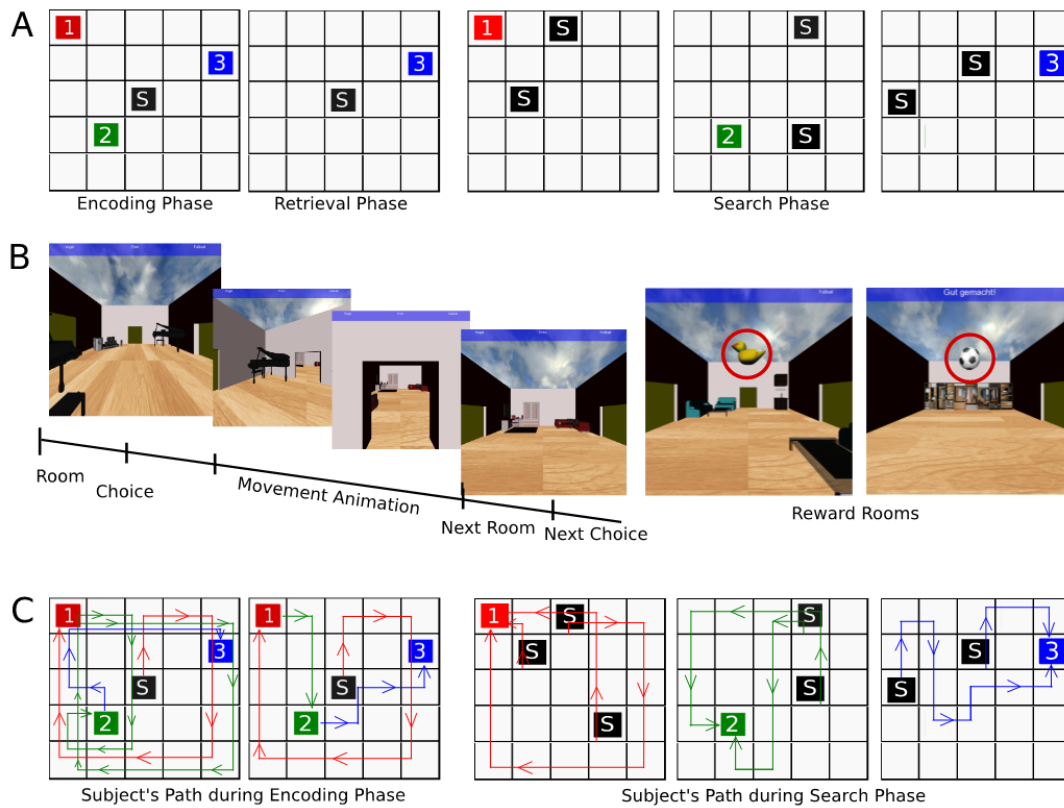
We then extracted  $\beta$ -parameter estimates using either model-free or model-based parametric value regressors in these ROIs and calculate the correlation between these  $\beta$  and the fitted parameter  $\omega$  from our behavioural hybrid learning model. This analysis provides additional information over the previous GLM analysis: the GLM identifies regions in which BOLD activity fluctuates with value signals of model-free and model-based RL on a population level without considering to which degree an individual subject of the group uses the corresponding strategy. The ROI analysis tests the hypothesis that activity in these regions, which correlate with value signals, is indeed related to the degree at which an individual subject employs that strategy behaviourally.

Since circular analysis, the use of the same dataset and contrast for selection and selective analysis, has been a common pitfall in systems neuroscience (Kriegeskorte et al., 2009),

we pay particular attention to avoid it here. Our ROI analysis is not subject to such ‘double-dipping’ for the following reasons:

- (1) We used a cross-validation leave-one-out procedure to define our regions of interest, ensuring that the data that is used to define the ROI is independent from the data extracted of this ROI.
- (2) The value-contrasts used in defining and extracting the ROI data ( $\beta$ ) are from a different RL model than the one providing the behavioural  $\omega$  parameter that is correlated against the ROI data, making both independent by design (see description above).

## 2.6 Figures

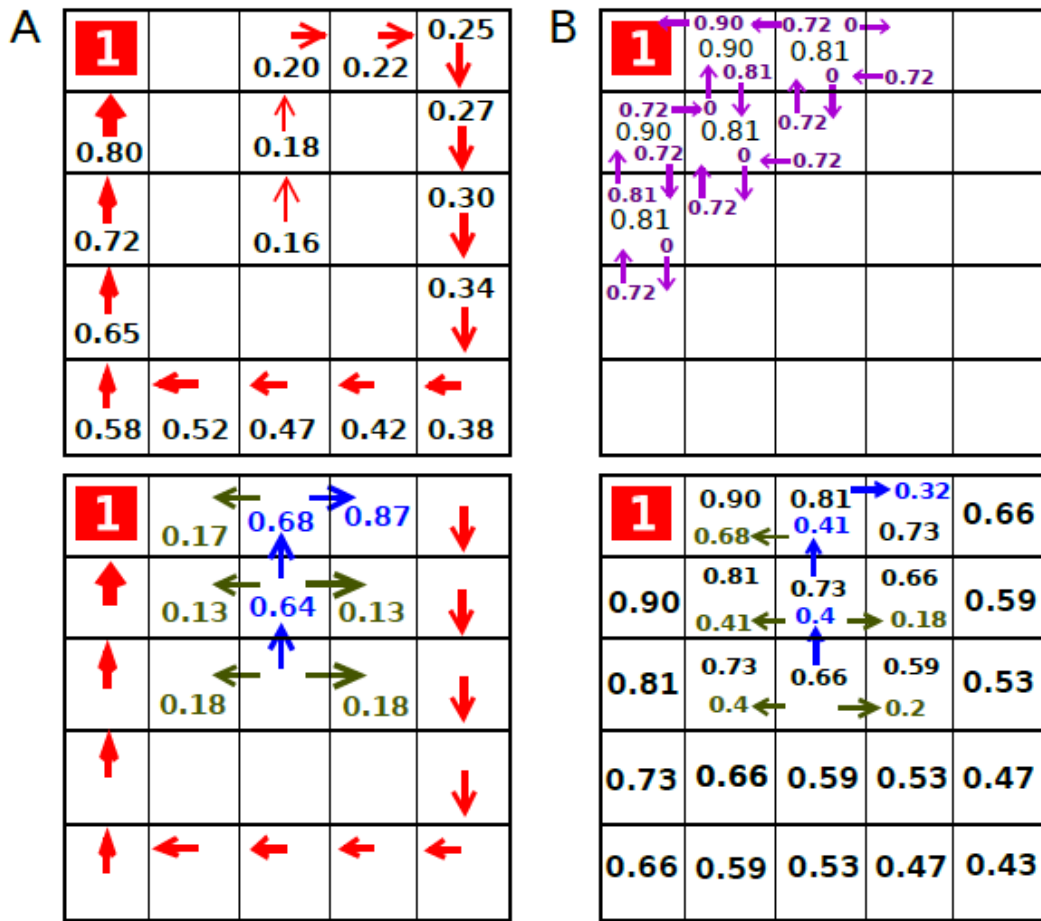


**Figure 2.1. Wayfinding task and behavioral results**

A. Layout of the grid-world. The Virtual Reality (VR) environment consisted of a 5 by 5 grid of rooms. Each square represents a room which contained distinct furniture and objects to distinguish individual rooms. Black square represents starting position, colored squares reward locations and the number represents the order in which they need to be found. The wayfinding task consisted of three phases: encoding, retrieval, and search phase. During the search phase, subject had to locate one randomly chosen reward at each trial, each time starting from a different starting position.

B. Screenshots of the virtual reality environment. Each room is furnished with distinct objects to allow subjects to distinguish and recognize individual rooms. At each room (decision point) subjects could choose up to three directions (corner rooms had either one or two directions to choose). After a choice was made, an animation was leading to the room in the selected direction; this movement lasted 2.5 – 3 seconds jittered uniformly. The next room and, if applicable, the reward were presented.

C. Path from a representative participant (subject no. 1) who exhibited a tendency towards route-based strategy. During the encoding phase, the subject established by repetition a fixed route from one reward to the other. During the search phase, the subject still followed the established route to reach the reward when it started from a location on that previous route. However, when the subject started from a position which was not part of the original route, she could locate the reward room using the shortest possible path.

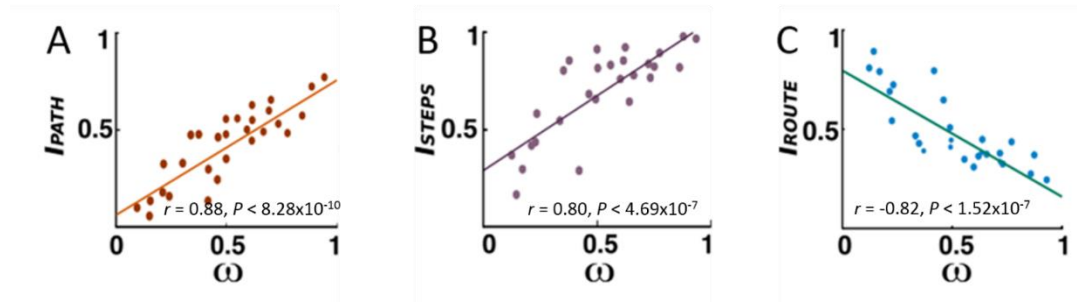


**Figure 2.2. Reinforcement learning models and model fits**

The top panel displays action values, showing how valuable it is to move along the route in a certain state; the bottom panel shows the probability of taking certain actions in those state based on the action values. Black numbers are state values, blue numbers are probabilities of chosen action, green values refer to probabilities of other not chosen actions. Note that not all probabilities for non-preferred actions are shown.

*A.* Model free valuation based on the SARSA ( $\lambda$ ) algorithm. After reaching a reward this algorithm updates the values only along the traversed path.

*B.* Model-based valuations derived from dynamic programming. The model-based algorithm updates values not only along the taken path, but across the entire grid world.



**Figure 2.3. Correlation between navigation indices and weight parameters (n=27)**

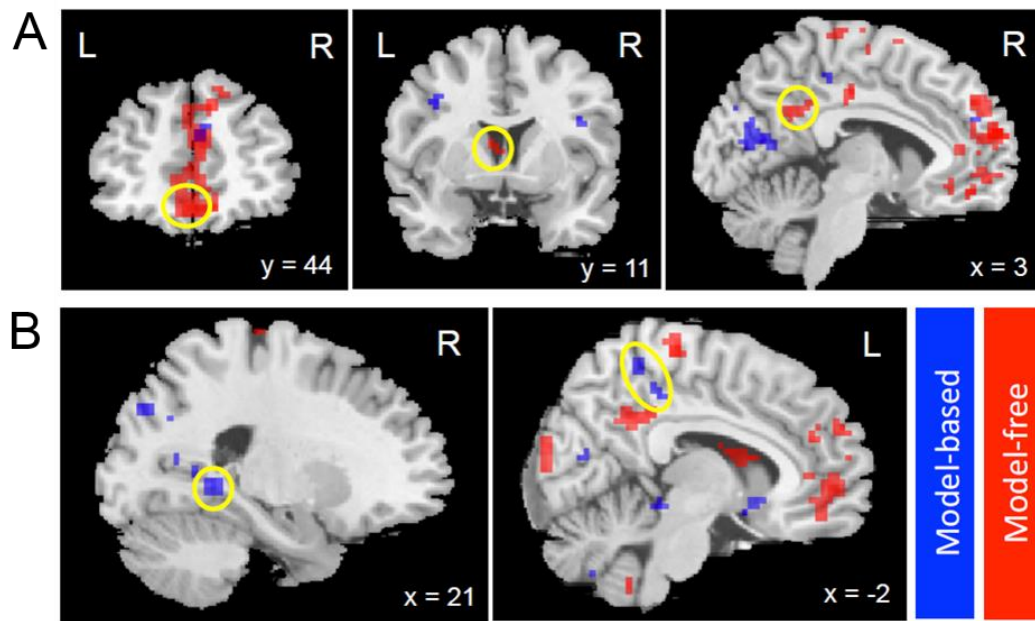
A. Significant positive correlation between  $I_{PATH}$  and  $\omega$  parameters.

B. Significant positive correlation between  $I_{STEP}$  and  $\omega$  parameters.

C. Significant negative correlation between  $I_{ROUTE}$  and  $\omega$  parameters for the fMRI experiment.

$I_{PATH}$ ,  $I_{STEP}$ ,  $I_{ROUTE}$ , and  $\omega$  are averaged values for individual subject across three different phases of the wayfinding task.



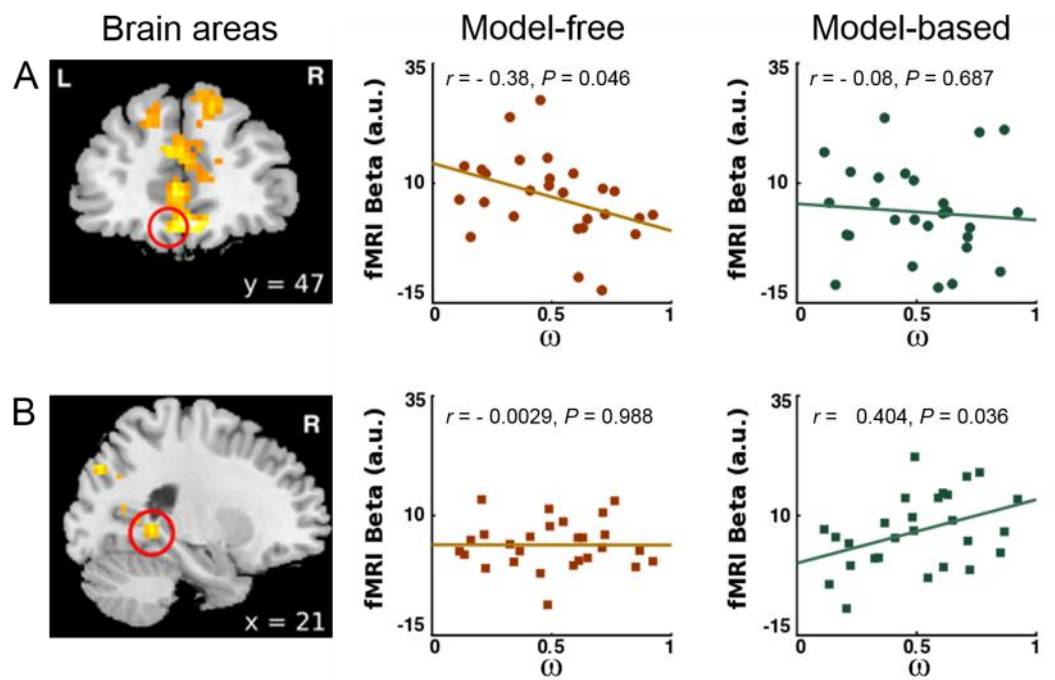


**Figure 2.4. Correlations of model predicted values with BOLD signals**

*A.* Correlates of model-free valuations in medial/vmPFC, striatum, and retrosplenial cortex.

*B.* Correlates of model-based valuation in parahippocampal and medial temporal lobe region as well as the left retrosplenial cortex.

Displayed results are significant at  $P < 0.05$  whole brain FWE corrected at the cluster level.



**Figure 2.5. Correlation between subjects' relative degree of model-based behavior and fMRI parameter estimates**

*A.* In vmPFC, we found a significant negative correlation between the  $\omega$  parameter in the behavioural hybrid model and  $\beta$ -estimates in the GLM from the parametric regressor of model-free values. That means, across subjects, the larger the relative degree of model-free choice behaviour of a subject, the stronger was her representation of model-free values in the BOLD signal in vmPFC. No such relationship was found for the model-based value regressor.

*B.* In the parahippocampal gyrus, we found a significant positive correlation between  $\omega$  and  $\beta$ - estimates from the parametric regressor of model-free values. The larger the relative degree of model-based choice behaviour in a subject, the bigger was her representation of model-based values in the BOLD signal in the parahippocampal gyrus. No such correlation was found for the model-free value regressor.

## 2.7 References

- Aguirre GK, D'Esposito M (1997) Environmental knowledge is subserved by separable dorsal/ventral neural areas. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 17:2512-2518.
- Aguirre GK, D'Esposito M (1999) Topographical disorientation: a synthesis and taxonomy. *Brain* 122 ( Pt 9):1613-1628.
- Apps MA, Ramnani N (2014) The anterior cingulate gyrus signals the net value of others' rewards. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 34:6190-6200.
- Balleine BW, O'Doherty JP (2010) Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* 35:48-69.
- Bar M, Aminoff E, Schacter DL (2008) Scenes unseen: the parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28:8539-8544.
- Basten U, Biele G, Heekeren HR, Fiebach CJ (2010) How the brain integrates costs and benefits during decision making. *Proc Natl Acad Sci U S A* 107:21767-21772.
- Boedecker J, Lampe T, Riedmiller M (2013) Modeling effects of intrinsic and extrinsic rewards on the competition between striatal learning systems. *Front Psychol* 4:739.
- Bohbot VD, Iaria G, Petrides M (2004) Hippocampal function and spatial memory: evidence from functional neuroimaging in healthy participants and performance of patients with medial temporal lobe resections. *Neuropsychology* 18:418-425.
- Brodt S, Pohlchen D, Flanagin VL, Glasauer S, Gais S, Schonauer M (2016) Rapid and independent memory formation in the parietal cortex. *Proc Natl Acad Sci U S A* 113:13251-13256.
- Burgess N, Maguire EA, O'Keefe J (2002) The human hippocampus and spatial and episodic memory. *Neuron* 35:625-641.
- Chersi F, Burgess N (2015) The Cognitive Architecture of Spatial Navigation: Hippocampal and Striatal Contributions. *Neuron* 88:64-77.

- Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci* 8:1704-1711.
- Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69:1204-1215.
- Doeller CF, King JA, Burgess N (2008) Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *Proc Natl Acad Sci U S A* 105:5915-5920.
- Doya K (1999) What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Netw* 12:961-974.
- Eichenbaum H, Cohen NJ (2014) Can we reconcile the declarative memory and spatial navigation views on hippocampal function? *Neuron* 83:764-770.
- Ekstrom AD, Arnold A, Iaria G (2014) A critical review of the allocentric spatial representation and its neural underpinnings: toward a network-based perspective. *Front Hum Neurosci* 8.
- Ekstrom AD, Kahana MJ, Caplan JB, Fields TA, Isham EA, Newman EL, Fried I (2003) Cellular networks underlying human spatial navigation. *Nature* 425:184-188.
- Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. *Nature* 392:598-601.
- Epstein RA, Parker WE, Feiler AM (2007) Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27:6141-6149.
- Etchamendy N, Bohbot VD (2007) Spontaneous navigational strategies and performance in the virtual town. *Hippocampus* 17:595-599.
- Ferbinteanu J (2016) Contributions of Hippocampus and Striatum to Memory-Guided Behavior Depend on Past Experience. *The Journal of Neuroscience* 36:6459-6470.
- Foo P, Warren WH, Duchon A, Tarr MJ (2005) Do humans integrate routes into a cognitive map? Map- versus landmark-based navigation of novel shortcuts. *J Exp Psychol Learn Mem Cogn* 31:195-215.

- Glascher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66:585-595.
- Glascher J, Adolphs R, Damasio H, Bechara A, Rudrauf D, Calamia M, Paul LK, Tranel D (2012) Lesion mapping of cognitive control and value-based decision making in the prefrontal cortex. *Proc Natl Acad Sci U S A* 109:14681-14686.
- Golledge RG (1999) *Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes*. Baltimore: The John Hopkins University Press.
- Hampton AN, Bossaerts P, O'Doherty JP (2006) The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 26:8360-8367.
- Hanel PH, Vione KC (2016) Do Student Samples Provide an Accurate Estimate of the General Public? *PloS one* 11:e0168354.
- Hartley T, Maguire EA, Spiers HJ, Burgess N (2003) The well-worn route and the path less travelled. Distinct neural bases of route-following and wayfinding in humans. *Neuron* 37:877-888.
- Hegarty M, Waller D (2005) Individual differences in spatial abilities. In: *Handbook of visuospatial thinking* (Shah P, Miyake A, eds), pp 121-169. New York: Cambridge University Press.
- Hunt LT, Kolling N, Soltani A, Woolrich MW, Rushworth MF, Behrens TE (2012) Mechanisms underlying cortical activity during value-guided choice. *Nat Neurosci* 15:470-476, s471-473.
- Iaria G, Petrides M, Dagher A, Pike B, Bohbot VD (2003) Cognitive strategies dependent on the hippocampus and caudate nucleus in human navigation: variability and change with practice. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 23:5945-5952.
- Iaria G, Chen JK, Guariglia C, Ptito A, Petrides M (2007) Retrosplenial and hippocampal brain regions in human navigation: complementary functional contributions to the formation and use of cognitive maps. *The European journal of neuroscience* 25:890-899.

- Igloi K, Zaoui M, Berthoz A, Rondi-Reig L (2009) Sequential egocentric strategy is acquired as early as allocentric strategy: Parallel acquisition of these two navigation strategies. *Hippocampus* 19:1199-1211.
- Iglói K, Doeller CF, Berthoz A, Rondi-Reig L, Burgess N (2010) Lateralized human hippocampal activity predicts navigation based on sequence or place memory. *Proceedings of the National Academy of Sciences* 107:14466-14471.
- Ino T, Inoue Y, Kage M, Hirose S, Kimura T, Fukuyama H (2002) Mental navigation in humans is processed in the anterior bank of the parieto-occipital sulcus. *Neurosci Lett* 322:182-186.
- Jacobs J, Weidemann CT, Miller JF, Solway A, Burke JF, Wei XX, Suthana N, Sperling MR, Sharan AD, Fried I, Kahana MJ (2013) Direct recordings of grid-like neuronal activity in human spatial navigation. *Nat Neurosci* 16:1188-1190.
- Janzen G, van Turenout M (2004) Selective neural representation of objects relevant for navigation. *Nat Neurosci* 7:673-677.
- Kahnt T, Park SQ, Haynes JD, Tobler PN (2014) Disentangling neural representations of value and salience in the human brain. *Proc Natl Acad Sci U S A* 111:5000-5005.
- Kaplan R, Bush D, Bonnefond M, Bandettini PA, Barnes GR, Doeller CF, Burgess N (2014) Medial prefrontal theta phase coupling during spatial memory retrieval. *Hippocampus* 24:656-665.
- King JA, Burgess N, Hartley T, Vargha-Khadem F, O'Keefe J (2002) Human hippocampus and viewpoint dependence in spatial memory. *Hippocampus* 12:811-820.
- Kolb B, Buhrmann K, McDonald R, Sutherland RJ (1994) Dissociation of the medial prefrontal, posterior parietal, and posterior temporal cortex for spatial navigation and recognition memory in the rat. *Cerebral cortex (New York, NY : 1991)* 4:664-680.
- Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* 12:535-540.
- Kurth-Nelson Z, Bickel W, Redish AD (2012) A theoretical account of cognitive effects in delay discounting. *The European journal of neuroscience* 35:1052-1064.

- Latini-Corazzini L, Nesa MP, Ceccaldi M, Guedj E, Thinus-Blanc C, Cauda F, Dagata F, Peruch P (2010) Route and survey processing of topographical memory during navigation. *Psychol Res* 74:545-559.
- Lin A, Adolphs R, Rangel A (2012) Social and monetary reward learning engage overlapping neural substrates. *Soc Cogn Affect Neurosci* 7:274-281.
- Maguire EA, Burgess N, O'Keefe J (1999) Human spatial navigation: cognitive maps, sexual dimorphism, and neural substrates. *Curr Opin Neurobiol* 9:171-177.
- Maguire EA, Burgess N, Donnett JG, Frackowiak RS, Frith CD, O'Keefe J (1998) Knowing where and getting there: a human navigation network. *Science* 280:921-924.
- Marchette SA, Bakker A, Shelton AL (2011) Cognitive mappers to creatures of habit: differential engagement of place and response learning mechanisms predicts human navigational behavior. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31:15264-15268.
- Marchette SA, Vass LK, Ryan J, Epstein RA (2014) Anchoring the neural compass: coding of local spatial reference frames in human medial parietal lobe. *Nat Neurosci* 17:1598-1606.
- Meder D, Kolling N, Verhagen L (2017) Simultaneous representation of a spectrum of dynamically changing value estimates during decision making. 8:1942.
- Miller AM, Vedder LC, Law LM, Smith DM (2014a) Cues, context, and long-term memory: the role of the retrosplenial cortex in spatial cognition. *Front Hum Neurosci* 8:586.
- Miller AMP, Vedder LC, Law LM, Smith DM (2014b) Cues, context, and long-term memory: the role of the retrosplenial cortex in spatial cognition. *Frontiers in Human Neuroscience* 8.
- Moscovitch M, Cabeza R, Winocur G, Nadel L (2016) Episodic Memory and Beyond: The Hippocampus and Neocortex in Transformation. *Annual review of psychology* 67:105-134.
- Moser EI, Kropff E, Moser MB (2008) Place cells, grid cells, and the brain's spatial representation system. *Annu Rev Neurosci* 31:69-89.

- Newman EL, Caplan JB, Kirschen MP, Korolev IO, Sekuler R, Kahana MJ (2007) Learning your way around town: How virtual taxicab drivers learn to use both layout and landmark information. *Cognition* 104:231-253.
- Nori R, Giusberti F (2006) Predicting cognitive styles from spatial abilities. *Am J Psychol* 119:67-86.
- O'Craven KM, Kanwisher N (2000) Mental Imagery of Faces and Places Activates Corresponding Stimulus-Specific Brain Regions. *J Cognitive Neuroscience* 12:1013-1023.
- O'Keefe J, Nadel L (1978) *The Hippocampus as A Cognitive Map*. Oxford: Clarendon Press.
- Pfeiffer BE, Foster DJ (2013) Hippocampal place cell sequences depict future paths to remembered goals. *Nature* 497:74-79.
- Redgrave P, Rodriguez M, Smith Y, Rodriguez-Oroz MC, Lehericy S, Bergman H, Agid Y, DeLong MR, Obeso JA (2010) Goal-directed and habitual control in the basal ganglia: implications for Parkinson's disease. *Nat Rev Neurosci* 11:760-772.
- Redish AD (1999) *Beyond the Cognitive Map*. Cambridge: MIT Press.
- Regier PS, Amemiya S, Redish AD (2015) Hippocampus and subregions of the dorsal striatum respond differently to a behavioral strategy change on a spatial navigation task. *Journal of neurophysiology* 114:1399-1416.
- Rosenbaum RS, Cassidy BN, Herdman KA (2015) Patterns of preserved and impaired spatial memory in a case of developmental amnesia. *Frontiers in Human Neuroscience* 9:196.
- Rosenbaum RS, Ziegler M, Winocur G, Grady CL, Moscovitch M (2004) "I have often walked down this street before": fMRI studies on the hippocampus and other structures during mental navigation of an old environment. *Hippocampus* 14:826-835.
- Rushworth MF, Behrens TE (2008) Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat Neurosci* 11:389-397.



- Shelton AL, Gabrieli JD (2002) Neural correlates of encoding space from route and survey perspectives. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 22:2711-2717.
- Shikauchi Y, Ishii S (2015) Decoding the view expectation during learned maze navigation from human fronto-parietal network. *Scientific Reports* 5:17648.
- Simon DA, Daw ND (2011) Neural correlates of forward planning in a spatial decision task in humans. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31:5526-5539.
- Spiers HJ, Barry C (2015) Neural systems supporting navigation. *Current Opinion in Behavioral Sciences* 1:47-55.
- Sutton RS, Barto AG (1998) Eligibility Traces. In: *Reinforcement Learning: An Introduction*. Cambridge: MIT Press.
- Tolman EC (1948) Cognitive map in rats and men. *The Psychological Review* 55:189-208.
- Urgolites ZJ, Kim S, Hopkins RO, Squire LR (2016) Map reading, navigating from maps, and the medial temporal lobe. *PNAS* 113:5.
- van der Meer MAA, Redish AD (2011) Ventral striatum: a critical look at models of learning and evaluation. *Current opinion in neurobiology* 21:387-392.
- Wiener JM, de Condappa O, Harris MA, Wolbers T (2013) Maladaptive bias for extrahippocampal navigation strategies in aging humans. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33:6012-6017.
- Wolbers T, Büchel C (2005) Dissociable retrosplenial and hippocampal contributions to successful formation of survey representations. *Journal of Neuroscience* 25:3333-3340.
- Wolbers T, Wiener JM (2014) Challenges for identifying the neural mechanisms that support spatial navigation: the impact of spatial scale. *Front Hum Neurosci* 8.
- Wolbers T, Weiller C, Büchel C (2004) Neural foundations of emerging route knowledge in complex spatial environments. *Cognitive Brain Research* 21:401-411.

- Wolbers T, Hegarty M, Buchel C, Loomis JM (2008) Spatial updating: how the brain keeps track of changing object locations during observer motion. *Nat Neurosci* 11:1223-1230.
- Wolbers T, Klatzky RL, Loomis JM, Wutte MG, Giudice NA (2011) Modality-independent coding of spatial layout in the human brain. *Curr Biol* 21:984-989.
- Wolbers T, Hegarty, M. (2010) What determines our navigational abilities? *Trends in Cognitive Sciences* 14:138-146.
- Wunderlich K, Rangel A, O'Doherty JP (2009) Neural computations underlying action-based decision making in the human brain. *Proc Natl Acad Sci U S A* 106:17199-17204.
- Wunderlich K, Dayan P, Dolan RJ (2011) Mapping value based planning and extensively trained choice in the human brain. *Nature Neuroscience* 15:786 - 791.
- Wunderlich K, Smittenaar P, Dolan RJ (2012) Dopamine enhances model-based over model-free choice behavior. *Neuron* 75:418-424.
- Zhang H, Copara M, Ekstrom AD (2012) Differential recruitment of brain networks following route and cartographic map learning of spatial environments. *PloS one* 7:e44886.

## 2.8 Supplementary Information

### 2.8.1 Supplementary Results and Discussion

#### ***Strategy Adoption in Three Different Phases of the Wayfinding Task***

To test if our RL algorithm captures the variability in strategy adoption during navigation we calculated correlation coefficients of both indices ( $I_{PATH}$  and  $I_{STEPS}$ ) with the fitted weight ( $\omega$ ) for each phase of the wayfinding task (Supplementary Figure 2.2).

During the encoding phase, we found significant correlation between  $\omega$  and  $I_{PATH}$  ( $r = 0.84$ , two-tailed t-test  $P < 5.8 \times 10^{-8}$ ),  $I_{STEPS}$  ( $r = 0.87$ , two-tailed t-test  $P < 7.6 \times 10^{-9}$ ), as well as  $I_{ROUTE}$  ( $r = -0.75$ , two-tailed t-test  $P < 8.7 \times 10^{-6}$ ). Similarly, in retrieval phase, we found significant correlation between  $\omega$  and  $I_{PATH}$  ( $r = 0.73$ , two-tailed t-test  $P < 1.4 \times 10^{-5}$ ),  $I_{STEPS}$  ( $r = 0.52$ , two-tailed t-test  $P < 0.006$ ), as well as  $I_{ROUTE}$  ( $r = -0.65$ , two-tailed t-test  $P < 2.6 \times 10^{-4}$ ). Lastly, in the search phase, we also found significant correlation between  $\omega$  and  $I_{PATH}$  ( $r = 0.56$ , two-tailed t-test  $P < 0.003$ ),  $I_{STEPS}$  ( $r = 0.40$ , two-tailed t-test  $P < 0.03$ ), as well as  $I_{ROUTE}$  ( $r = -0.45$ , two-tailed t-test  $P < 0.02$ ). For aggregated data over all experimental phases see main text.

#### ***Additional Behavioural Experiment without Monetary Reward***

To rule out the influence of an external monetary reward on our results, sixteen healthy female participants (19 – 29 years of age) took part in our additional behavioural experiment. The experimental procedure was almost the same as the main experiment, except subjects completed 12 trials (instead of as many as possible within a certain time) during the search phase. For this additional behavioural experiment, no reward is associated with target locations. In addition, subjects received a flat participation credit instead of monetary rewards for their participation.

We compared the fitted parameters from this behavioural experiment to the ones from the fMRI experiment. Comparing the fitted parameters from the additional experiment (without reward) and the original experiment (with reward) using an independent-samples t-test, we found no significant differences for all parameter values. Model free ( $\alpha$ :  $t(41) = 0.44$ ,  $P > 0.66$ ;  $\lambda$ :  $t(41) = 0.61$ ,  $P > 0.54$ ;  $\beta$ :  $t(41) = -0.78$ ,  $P > 0.430$ ), model-based ( $\beta$ :  $t(41) = 0.52$ ,  $P > 0.60$ ), and hybrid model ( $\alpha$ :  $t(41) = 0.79$ ,  $P > 0.22$ ;  $\lambda$ :  $t(41) =$

-0.58,  $P > 0.56$ ;  $\beta$ :  $t(41) = 1.72$ ,  $P > 0.09$ ; and  $\omega$ :  $t(41) = -1.62$ ,  $P = 0.11$ ). In addition, we also replicate the correlation between parameter weight ( $\omega$ ) from the hybrid model and navigation indices. For the encoding phase, we found significant positive correlation between  $\omega$  and  $I_{PATH}$  ( $r = 0.54$ , two-tailed t-test  $P < 0.03$ ) and  $I_{STEPS}$  ( $r = 0.61$ , two-tailed t-test  $P < 0.01$ ). For the retrieval phase, we found significant correlation between  $\omega$  and  $I_{PATH}$  ( $r = 0.91$ , two-tailed t-test  $P < 4.7 \times 10^{-7}$ ),  $I_{STEPS}$  ( $r = 0.60$ , two-tailed t-test  $P < 0.01$ ), as well as  $I_{ROUTE}$  ( $r = -0.822$ , two-tailed t-test  $P < 9.3 \times 10^{-5}$ ). We also found significant correlation between  $\omega$  and  $I_{PATH}$  ( $r = 0.83$ , two-tailed t-test  $P < 6.6 \times 10^{-5}$ ),  $I_{STEPS}$  ( $r = 0.72$ , two-tailed t-test  $P < 0.001$ ), as well as  $I_{ROUTE}$  ( $r = -0.78$ , two-tailed t-test  $P < 2.7 \times 10^{-4}$ ) for the search phase.

### ***Additional exploratory hypotheses tested in the fMRI analysis***

We tested a few exploratory hypotheses, in addition to the results presented in the main text, and mention the results here for completeness.

First, there are several cognitive processes that subjects are faced with when entering each room, such as recognition of the objects, recall and matching to previously encountered rooms, identifying the location of the room in a cognitive map, calculating a navigation strategy and finally deriving an action choice from that. A lot of these processes have to take place before any decision can be made. Since our 3D environment is visually much more complex than a decision experiment where simple stimuli are presented, we expected the time of the response to be a better predictor than entering the room for when the brain computes navigation decisions. We therefore hypothesized that value signals would be represented strongest at the time of making a choice, which is the time information that we used in our GLM. However, we alternatively explored if there is any BOLD activity pertaining to model-free or model-based values already when subjects entered the room, i.e. preceding the actual decision point. From this analysis, we found neural correlates of model-free values in the right precuneus [ $x = 3$ ,  $y = -70$ ,  $z = 38$ ] and of model-based values in right middle occipital gyrus [ $x = 39$ ,  $y = -79$ ,  $z = 14$ ] but as expected those relationships were much weaker than when they are related to the time of the choice.

We also tested if the brain tracks the frequency of entering the current room, frequency of entering the current room and choosing the same direction, time since last entering

the room, as well as whether the current room was located in the centre or the peripheral of the grid-world. We did not find any BOLD activity that are significantly correlated with these variables.

We also tested the number of trials and MRT scores as covariates in our 2<sup>nd</sup> level GLM analysis. There was no significant correlation between these covariates and BOLD signal anywhere in the brain that survived correction for multiple comparisons.

Lastly, since putting model-free and model-based regressors within one GLM might remove potentially important shared variance, we also created separate GLMs with model-free or model-based parametric modulators. In line with our main results, BOLD activity in medial prefrontal, striatal, and retrosplenial regions pertain to model-free valuations while BOLD activity in parahippocampal, hippocampal, and retrosplenial regions pertain to model-based valuations. We did not, however, find any differences in activity pattern across three different phases of the wayfinding task.

## **2.8.2 Supplementary Methods**

### ***Wayfinding Task***

We built two 3D virtual environments with a commercially available software that displays a first person point of view in a fully textured grid world (Vizard version 4.0; WorldViz, LLC <https://worldviz.com>). The 3D environment consisted of a 5 by 5 grid of rooms. Every room contained distinct furniture and objects as landmarks to distinguish individual rooms. Three rooms were designated as reward rooms. Each reward room contained one of three objects that subjects had to find while navigating through the maze. For the practice session, conducted three to five days before the fMRI scanning sessions, subject practiced on a smaller 4 by 4 grid world with different room furnishings but otherwise similar to the main task. Subjects could move from one room to another by pressing keys on a button box. Subjects could move forward, left or right. Backtracking was not allowed. Following their response, subjects viewed an animated movement sequence of the selected direction. It is important to note that a wayfinding task where participants are required to choose a direction and are then moved through

the environment based on their directional choices is well established in the field of spatial navigation (for more examples see the following references: (Harris et al., 2012; Latini-Corazzini et al., 2010; Shikauchi and Ishii (2015); Viard et al., 2011; Voermans et al., 2004, Wiener et al., 2013)).

After entering the MRI, we instructed subjects to freely explore the 5 by 5 grid, without any rewards present, to learn to move with the button press and to get comfortable with the stimulus materials. Subjects had 35 choices to explore the setting which gave them enough time to explore all possible states of the environment, i.e. all the rooms, during this initial exploration. The following task consisted of three phases: (1) encoding, (2) retrieval, and (3) search. During the encoding phase, subjects always started in the same starting position and were asked to collect three rewards in a specific order over eight trials. During the retrieval phase, subjects were instructed to collect one specific reward at a time from the same starting position as in the encoding phase. Rewards were positioned at the same location as in encoding phase and selected in random order. Subject completed 15 trials in the retrieval phase. During the search phase, subjects had to collect a specific randomly chosen reward, each time starting from a different starting position. Subjects had 20 minutes to do as many trials as possible. Note that using short routes would allow subjects to reach target rooms more quickly and hence collect more rewards. After the experiment subjects a part of subjects' financial compensation was related to the number of collected rewards. Use of a map-based strategy in this phase was therefore beneficial to subjects' payment.

### ***Mental Rotation Task***

The Vandenberg and Kuse Mental Rotation Test (MRT) is comprised of 24 items, six items on four separate pages in the test booklet (Vandenberg and Kuse, 1978). Each item is comprised of a row of five line drawings including a geometrical target figure in the left-most position followed by four response-choice figures: two rotated reproductions of the target and two distractors. The subject's task is to indicate which two of the four response choice figures are rotated reproductions of the target figure. In each item there are always two and only two correct figures and two incorrect distractor figures. For each item, subjects were instructed to find two response choices with figures identical to the target figures. Subjects had 10 minutes to complete the task and were informed when there were 5 minutes remaining and again when there were 2 minutes

remaining. Instructions emphasized that subjects should refrain from guessing. Score of 1 is given for every correct answer. Thus, the maximum score of the MRT is 48 (Caissie et al., 2009).

### **Model Fitting**

For each algorithm, we calculated a probability of choice based on the following softmax decision rule (Luce choice rule):

$$p_t = \frac{\exp \beta Q_t(s_t, a)}{\sum_{a' \in A} \exp \beta Q_t(s_t, a')} \quad [\text{S2.1}]$$

The parameter  $\beta$  is the inverse temperature representing the degree of stochasticity of subjects' action selection.

We estimated a set of free parameters ( $\theta$ ) for each algorithm separately for each subject by mean of hierarchical model fitting (Wunderlich et al., 2012). First we applied logistic ( $\alpha, \omega, \lambda$ ) and exponential ( $\beta$ ) transformation before fitting parameters to transform bounded parameter into Gaussian distributed parameter values with population mean and standard deviation. In the equation below, the Greek alphabets represent the parameters we used in the model, while the Latin alphabets stand for parameters in the logistic transformations that range from  $-\infty$  to  $\infty$ .

$$\alpha = \frac{1}{1+\exp(-a)}; \quad \omega = \frac{1}{1+\exp(-w)}; \quad \lambda = \frac{1}{1+\exp(-l)}; \quad \beta = \exp(b) \quad [\text{S2.2}]$$

Next, we fitted these transformed parameters by maximizing the likelihood of all observed choices given the parameterized model:

$$L = P(c_i | \mu_a, \mu_l, \mu_b, \mu_w, \sigma_a, \sigma_l, \sigma_b, \sigma_w) = \int da_i dl_i db_i dw_i (c_i | a_i, l_i, b_i, w_i) (a_i | \mu_a, \sigma_a) P(l_i | \mu_l, \sigma_l) P(b_i | \mu_b, \sigma_b) P(w_i | \mu_w, \sigma_w) \quad [\text{S2.3}]$$

We later estimated mean and variance of the parameter distribution in the population based on our subject sample. As an example, for parameter  $\alpha$ :

$$\mu_a = \frac{1}{N} \sum_i^N a_i \quad \text{and} \quad \sigma_a = \sqrt{\frac{1}{N} \sum_i^N (a_i - \mu_a)^2} \quad [\text{S2.4}]$$

Following mean and variance estimation, we refitted single subject parameter values by minimizing over both the negative log likelihood of subjects' choice given the parameters and the negative log likelihood for individual subject parameter values given the distribution of parameters in the population:

$$P(a_i, l_i, b_i, w_i | c_i, \mu_a, \mu_l, \mu_b, \mu_w, \sigma_a \sigma_l, \sigma_b, \sigma_w) \\ \propto (c_i | a_i l_i b_i w_i) \times (a_i, l_i, b_i, w_i | \mu_a, \mu_l, \mu_b, \mu_w, \sigma_a \sigma_l, \sigma_b, \sigma_w) \quad [\text{S2.5}]$$

Estimation of the free parameters was implemented in Matlab using the Global Optimization Toolbox.

### ***Model evidence***

To approximate the model evidence, we computed the Bayesian Information criterion (BIC) as follows:

$$l(\hat{\theta}) + \frac{m}{2} \log n \quad [\text{S2.6}]$$

where  $l(\theta)$  is the negative log-likelihood of data at the maximum likelihood parameters  $\theta$ ;  $m$  is the number of free parameters optimized; and  $n$  is the number of choices the subject made. The BIC was calculated based on the hierarchical model fitting. Note that since subjects might employ either model-free or model-based strategy in individual trials, we did not choose RL model based on aggregated BIC of the subjects.

### ***Magnetic Resonance Imaging (MRI) Images Acquisition***

We performed functional imaging using a 3T whole-body Siemens MAGNETOM Verio scanner with an 8-channel head coil located at the Klinikum der Universität München. T2\* echo-planar images were obtained with 2390 ms repetition time (TR) with an acquisition matrix of  $64 \times 64$ , an echo time (TE) of 30 ms, a flip angle of  $90^\circ$ , and field of view of  $192 \times 192$  mm. Each volume consisted of 30 ( $3.0 \times 3.0 \times 3.0$  mm<sup>3</sup> voxels) axial slices with 15% gap. We completed five dummy scans at the beginning of each run



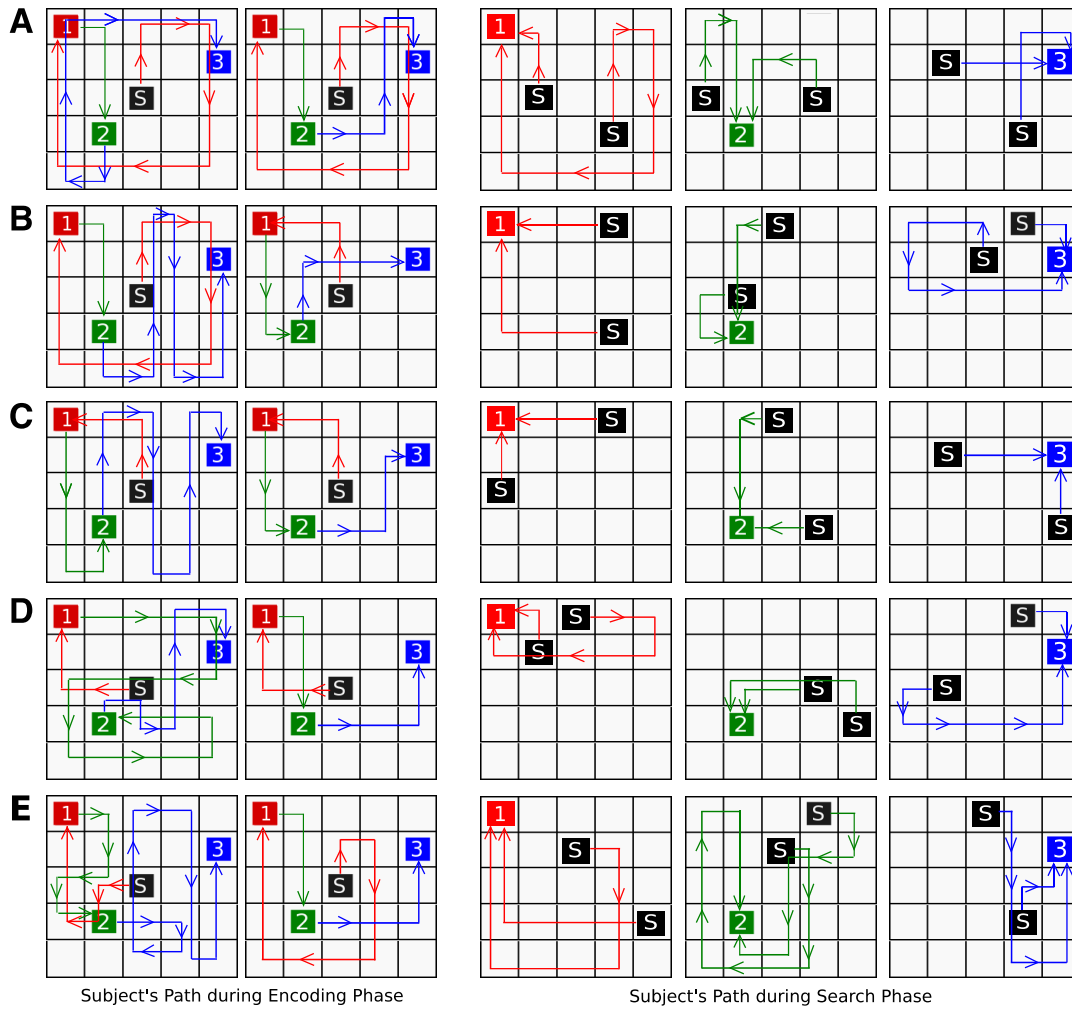
to allow for stabilization of the MR signal. In addition, a structural whole brain scan was acquired using an MP-RAGE T1-weighted sequence (TR/TE = 11.0 / 4.76 ms) with  $256 \times 256 \times 160$  acquisition matrix,  $1 \times 1 \times 1$  mm<sup>3</sup> voxel size, 15° flip angle, and a field of view of  $256 \times 256$  mm.

### ***Preprocessing of fMRI data***

Neuroimaging data were processed and analysed using SPM12 toolbox ([www.fil.ion.ucl.ac.uk/spm/software/spm12/](http://www.fil.ion.ucl.ac.uk/spm/software/spm12/)). An event-related statistical analysis was applied to the images on two levels using the general linear model approach as implemented in SPM12.

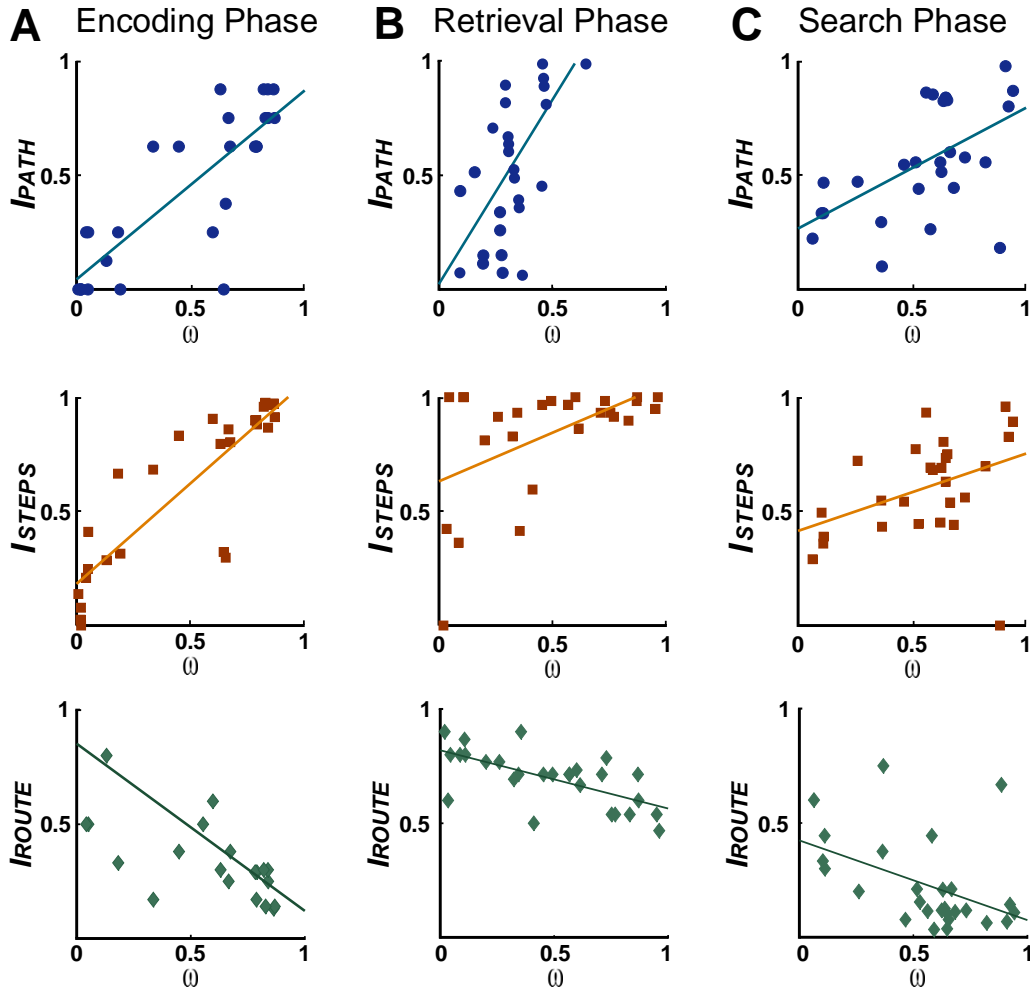
Functional images were realigned for head motion and coregistered between runs and to the structural images. The images were then spatially normalized to Montreal Neurological Institute (MNI) space using the normalization parameters generated during the segmentation for each subject's anatomical T1 scan and resampled to 2 mm isotropic voxels. Subsequently, all images were smoothed with an 8 mm full width at half maximum (FWHM) Gaussian kernel.

2.8.3 Supplementary Figures



**Supplementary Figure 2.1. Example of five representative subjects' paths in encoding and search phases**

During the encoding phase, most subjects started by exploring the environment. They then either: *A.* established a certain route to go from one reward to the next or *B.* found the shortest paths to go between one reward and the next. During retrieval and search phases some subjects mostly used the *C.* shortest path to retrieve rewards while others either *D.* took a detour before reaching the reward or *E.* used the route they established during the earlier encoding phase. No subject chose exclusively one strategy over the other. Note that starting direction was always facing upward and subjects could not go backwards, so that from starting position S the downward path was not permitted



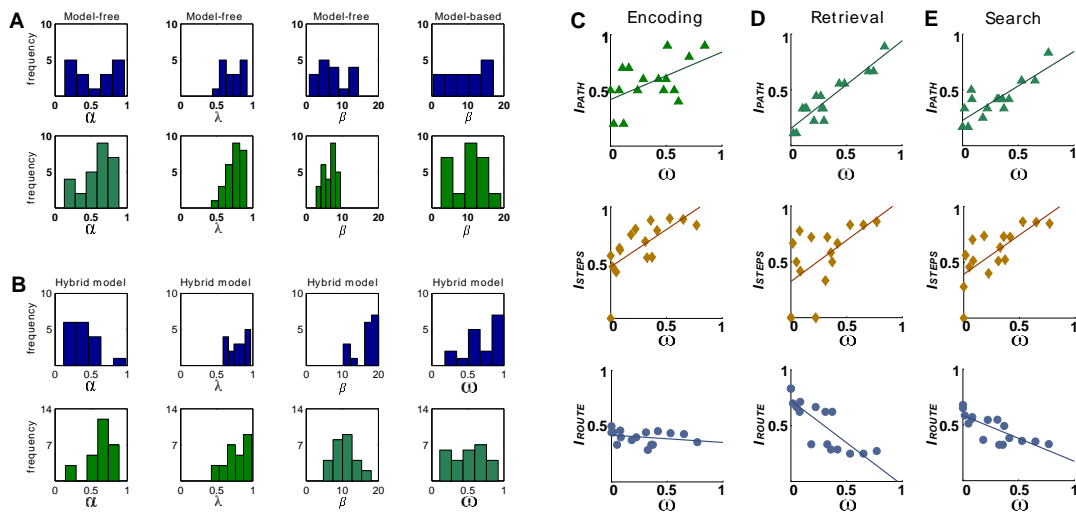
**Supplementary Figure 2.2. Correlation of navigation indices with parameter weight on every phase of the wayfinding task**

*A.* For the encoding phase, significant correlation between  $\omega$  and  $I_{PATH}$  ( $r = 0.84$ , two-tailed t-test  $P < 5.8 \times 10^{-8}$ ),  $I_{STEPS}$  ( $r = 0.87$ , two-tailed t-test  $P < 7.6 \times 10^{-9}$ ), as well as  $I_{ROUTE}$  ( $r = -0.75$ , two-tailed t-test  $P < 8.7 \times 10^{-6}$ ).

*B.* For the retrieval phase, significant correlation between  $\omega$  and  $I_{PATH}$  ( $r = 0.73$ , two-tailed t-test  $P < 1.4 \times 10^{-5}$ ),  $I_{STEPS}$  ( $r = 0.52$ , two-tailed t-test  $P < 0.006$ ), as well as  $I_{ROUTE}$  ( $r = -0.65$ , two-tailed t-test  $P < 2.6 \times 10^{-4}$ ).

*C.* For the search phase, significant correlation between  $\omega$  and  $I_{PATH}$  ( $r = 0.56$ , two-tailed t-test  $P < 0.003$ ),  $I_{STEPS}$  ( $r = 0.40$ , two-tailed t-test  $P < 0.03$ ), as well as  $I_{ROUTE}$  ( $r = -0.45$ , two-tailed t-test  $P < 0.02$ ).

These correlations confirm that RL algorithms capture the variability in strategy adoption during navigation.

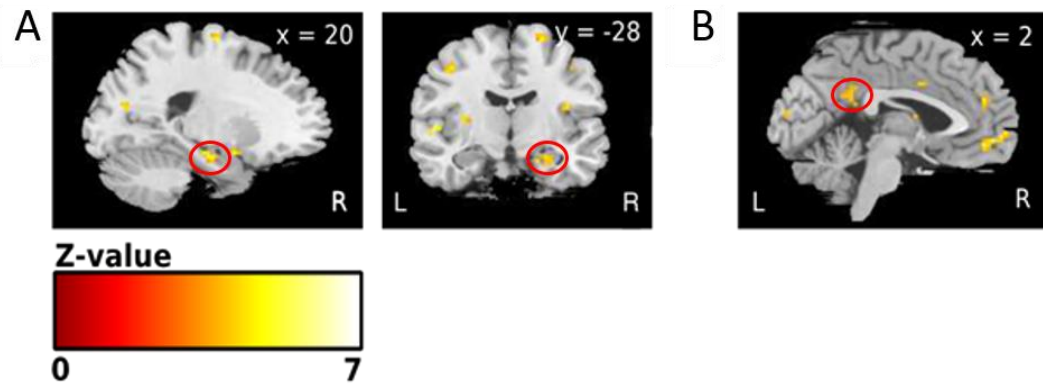


**Supplementary Figure 2.3. Distribution of fitted parameters for the behavioural and fMRI experiments**

*A.* Distribution of fitted parameters from the additional behavioural experiment and the fMRI experiment for  $\alpha$ ,  $\lambda$ , and  $\beta$  from model free and parameter  $\beta$  from model-based. We found no significant difference between fitted parameters from the behavioural and fMRI experiment for model-free ( $\alpha$ :  $P > 0.660$ ;  $\lambda$ :  $P > 0.546$ ;  $\beta$ :  $P > 0.430$ ), model-based ( $\beta$ :  $P > 0.60$ ).

*B.* Distribution of fitted parameters from the additional behavioural experiment and the fMRI experiment for  $\alpha$ ,  $\lambda$ ,  $\beta$ , and  $\omega$  from hybrid model. We found no significant difference between fitted parameters from the behavioural and fMRI experiment for hybrid model ( $\alpha$ :  $P > 0.216$ ;  $\lambda$ :  $P > 0.560$ ;  $\beta$ :  $P > 0.067$ ; and  $\omega$ :  $P > 0.12$ ).

*C.* For the encoding phase, significant correlation between  $\omega$  and IPATH ( $r = 0.54$ , two-tailed t-test  $P < 0.03$ ) and ISTEPS ( $r = 0.61$ , two-tailed t-test  $P < 0.01$ ). (D) For the retrieval phase, significant correlation between  $\omega$  and IPATH ( $r = 0.91$ , two-tailed t-test  $P < 4.7 \times 10^{-7}$ ), ISTEPS ( $r = 0.60$ , two-tailed t-test  $P < 0.01$ ), as well as IROUTE ( $r = -0.822$ , two-tailed t-test  $P < 9.3 \times 10^{-5}$ ). (E) For the search phase, significant correlation between  $\omega$  and IPATH ( $r = 0.83$ , two-tailed t-test  $P < 6.6 \times 10^{-5}$ ), ISTEPS ( $r = 0.72$ , two-tailed t-test  $P < 0.001$ ), as well as IROUTE ( $r = -0.78$ , two-tailed t-test  $P < 2.7 \times 10^{-4}$ ).



Supplementary Figure 2.4. Distance and model-based regressor in one GLM

Neural correlates of model-based regressor in *A*. parahippocampal/hippocampal area and *B*. retrosplenial cortex. Even when we put model-based and distance regressors as non-orthogonalized parametric modulators in the same GLM, we still see the correlated BOLD activity to model-based valuations. Note that the model-based algorithm computed value regressors using both subject's state (i.e. location relative to goal) and chosen action. Consequently, for some trials, model-based regressors did not necessarily reflect the optimal value of the subject's state. These findings rule out that the BOLD correlations with model-based value signals are only a spurious correlate of the distance to the reward.

## 2.8.4 Supplementary Tables

Supplementary Table 2.1. Navigation Indices, Fitted Parameter Weight ( $\omega$ ), and MRT Score

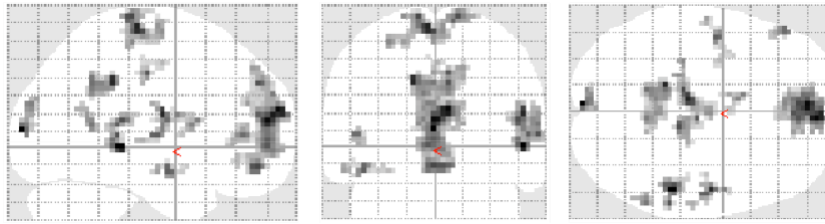
Subject	$I_{PATH}$	$I_{STEPS}$	$I_{ROUTE}$	Weight ( $\omega$ )				Aggregate BIC			MRT
				Encoding	Retrieval	Search	Average	MF	MB	Hybrid	
1	0.12	0.36	0.80	0.19	0.09	0.06	0.11	321.2	359.3	337.1	14
2	0.46	0.84	0.38	0.18	0.34	0.56	0.36	205.0	206.0	214.5	33
3	0.35	0.65	0.49	0.05	0.73	0.67	0.48	319.7	321.1	329.6	31
4	0.06	0.15	0.88	0.02	0.02	0.37	0.13	520.9	543.3	536.4	7
5	0.76	0.96	0.23	0.86	0.96	0.94	0.92	198.8	130.3	168.6	36
6	0.48	0.88	0.43	0.60	0.77	0.92	0.76	228.7	175.6	214.2	18
7	0.31	0.42	0.53	0.13	0.41	0.11	0.22	445.7	440.0	457.6	9
8	0.62	0.83	0.33	0.84	0.71	0.59	0.71	173.1	176.4	174.5	17
9	0.58	0.83	0.37	0.67	0.57	0.65	0.63	219.4	204.3	229.0	37
10	0.45	0.63	0.44	0.55	0.53	0.78	0.61	223.0	206.9	227.7	23
11	0.53	0.75	0.32	0.79	0.75	0.62	0.72	260.8	227.3	255.9	21
12	0.49	0.77	0.36	0.67	0.60	0.68	0.65	242.4	211.5	238.8	36
13	0.58	0.81	0.26	0.87	0.95	0.73	0.85	260.5	219.0	247.3	26
14	0.25	0.67	0.64	0.64	0.45	0.26	0.45	337.6	332.7	349.3	19
15	0.55	0.90	0.41	0.78	0.05	0.63	0.49	170.0	157.8	167.3	32
16	0.15	0.28	0.78	0.02	0.32	0.89	0.41	527.1	522.4	546.1	16
17	0.33	0.53	0.46	0.04	0.83	0.10	0.33	407.1	393.6	417.9	31
18	0.48	0.80	0.44	0.84	0.11	0.53	0.49	189.1	178.3	194.2	36
19	0.18	0.42	0.68	0.01	0.03	0.58	0.21	497.7	499.9	513.6	20
20	0.55	0.82	0.34	0.79	0.20	0.66	0.55	225.6	214.1	223.5	35
21	0.48	0.75	0.29	0.34	0.61	0.82	0.59	281.3	259.3	280.1	33
22	0.46	0.79	0.42	0.45	0.11	0.47	0.34	220.5	215.5	226.1	34
23	0.71	0.97	0.36	0.82	0.87	0.91	0.87	175.2	115.6	153.8	38
24	0.16	0.57	0.72	0.05	0.26	0.37	0.23	399.8	393.5	413.8	11
25	0.13	0.28	0.78	0.02	0.35	0.11	0.16	310.5	398.0	326.8	18
26	0.60	0.83	0.37	0.63	0.87	0.63	0.71	223.2	202.3	220.6	34
27	0.52	0.91	0.35	0.83	0.49	0.52	0.61	191.7	156.7	183.5	27

Supplementary Table 2.2. Distribution of subjects' individual maximum likelihoods and parameter estimates

	Model-free		Model-based		Hybrid		Random					
<b>Encoding Phase</b>												
$\beta$	5.70	<b>11.06</b>	13.43	3.01	<b>10.39</b>	19.54	11.61	<b>13.93</b>	17.67			
$\gamma$				-	<b>0.1</b>	-	-	<b>0.1</b>	-			
$\alpha$	0.08	<b>0.36</b>	0.62				0.09	<b>0.23</b>	0.76			
$\lambda$	0.67	<b>0.78</b>	0.90				0.74	<b>0.89</b>	0.95			
$\omega$							0.08	<b>0.59</b>	0.78			
NLL	79.13	<b>95.63</b>	210.32	68.81	<b>94.42</b>	232.85	68.04	<b>91.53</b>	210.22	116.29	<b>129.74</b>	238.42
BIC	86.39	<b>103.19</b>	218.78	71.26	<b>96.94</b>	235.69	78.86	<b>138.24</b>	222.52	116.29	<b>129.74</b>	238.42
<b>Retrieval Phase</b>												
$\beta$	5.08	<b>6.16</b>	9.20	9.79	<b>13.49</b>	19.06	8.63	<b>11.17</b>	15.34			
$\gamma$				-	<b>0.1</b>	-	-	<b>0.1</b>	-			
$\alpha$	0.32	<b>0.89</b>	0.96				0.58	<b>0.97</b>	0.99			
$\lambda$	0.55	<b>0.88</b>	0.94				0.41	<b>0.83</b>	0.97			
$\omega$							0.22	<b>0.49</b>	0.75			
NLL	20.15	<b>27.25</b>	43.21	17.43	<b>27.20</b>	41.82	17.23	<b>27.03</b>	41.24	51.39	<b>55.44</b>	65.70
BIC	25.92	<b>33.20</b>	49.31	19.36	<b>29.19</b>	43.91	25.93	<b>35.97</b>	50.41	51.39	<b>55.44</b>	65.70
<b>Search Phase</b>												
$\beta$	1.89	<b>2.49</b>	3.53	3.43	<b>6.55</b>	9.05	2.89	<b>5.05</b>	7.15			
$\gamma$							-	<b>0.1</b>	-			
$\alpha$	0.38	<b>0.87</b>	0.97				0.56	<b>0.87</b>	0.97			
$\lambda$	0.6	<b>0.86</b>	0.86				0.63	<b>0.86</b>	0.97			
$\omega$							0.39	<b>0.62</b>	0.72			
NLL	78.29	<b>96.19</b>	117.23	77.22	<b>94.69</b>	118.74	76.14	<b>96.13</b>	116.81	116.65	<b>124.48</b>	132.06
BIC	85.87	<b>103.69</b>	124.77	79.59	<b>97.20</b>	121.31	85.92	<b>106.13</b>	126.86	116.65	<b>124.48</b>	132.06
$\rho^2$	0.17	<b>0.29</b>	0.35	0.11	<b>0.31</b>	0.38	0.16	0.32	0.40			

Quartiles (median in bold) of best fitting parameters for the three algorithms used to produced regressors for the imaging analysis, along with the negative log likelihood (NLL), BIC estimated evidence, and pseudo- $r^2$  measure of individual fit.

Supplementary Table 2.3. Activation at decision points for model-free valuation across three phases of the wayfinding task

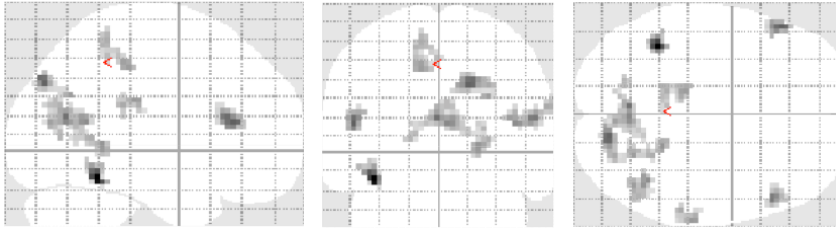


Regions	Peak Coordinates			Hemi	$T_{PEAK}$	Z score	k
	x	y	z				
Medial PFC							
Superior medial gyrus	6	59	20	R	6.29	4.86	524
Middle orbital gyrus	-3	47	-13	L	5.43	4.83	524
Anterior Cingulate Cortex	-3	38	5	L	4.58	4.40	524
Precentral gyrus							
	-6	-25	68	L	5.92	4.67	120
	-30	-28	59	L	4.59	3.89	120
	12	-31	74	R	5.33	4.34	49
Caudate Nucleus	-9	11	14	L	4.69	4.11	29
Middle temporal gyrus							
	51	-34	-1	R	6.47	4.95	73
	-45	-7	-16	L	5.10	4.21	35
Superior temporal gyrus							
	51	-10	17	R	5.86	4.64	52
	60	-22	2	R	4.09	4.54	52
Retrosplenial cortex (RSC)							
	-15	-40	35	L	5.85	4.63	130
	6	-52	32	R	4.72	3.98	130
	6	-19	41	R	5.22	4.28	37
Cuneus	0	-85	26	L	4.28	3.84	63
Calcarine gyrus	-3	-94	8	L	6.49	4.94	63

Regions showing correlated BOLD activity with model-free values as parametric regressors across three different phases of the wayfinding task. Statistical significance was determined at the group level using a random-effect analysis. Regions listed exhibited significant peak voxels at probability threshold of  $P < 0.05$  based on a FWE cluster level small volume correction ( $k = 25$ ). Peak voxel MNI coordinates x, y, z are given in millimeters. L, left; R, right.



Supplementary Table 2.4. Activation at decision points for model-based valuation across three phases of the wayfinding task



Regions	Coordinates			Hemi	TPEAK	Z score	k
	x	y	z				
Lateral PFC							
Inferior frontal gyrus	-51	29	14	L	5.29	4.32	97
Inferior frontal gyrus	54	23	17	R	4.93	4.11	35
Cuneus	15	-79	35	R	5.26	4.30	61
Precuneus	27	-64	29	R	4.39	3.77	61
Supramarginal gyrus	66	-28	26	R	4.51	3.84	30
Middle temporal gyrus	42	-55	14	R	4.69	3.95	43
Fusiform gyrus	-39	-49	-16	L	6.85	5.13	36
PPA							
Parahippocampal gyrus	21	-46	2	R	4.54	3.85	153
Calcarine gyrus	-3	-67	17	L	5.08	4.20	153
	5	-75	14	R	4.38	3.75	153
Retrosplinal cortex (RSC)	-12	-31	44	L	4.39	3.75	56

Regions showing correlated BOLD activity with model-based values as parametric regressors across three phases of the wayfinding task. Statistical significance was determined at the group level using a random-effect analysis. Regions listed exhibited significant peak voxels at probability threshold of  $P < 0.05$  based on a FWE cluster level small volume correction ( $k = 25$ ). Peak voxel MNI coordinates x, y, z are given in millimeters. L, left; R, right.

### 2.8.5 Supplementary References

- Caissie, A. F., Vigneau, F. & Bors, D. A. What does the Mental Rotation Test Measure? An Analysis of Item Difficulty and Item Characteristics. *The Open Psychology Journal* 2, 8 (2009).
- Harris, M. A., Wiener, J. M. & Wolbers, T. Aging specifically impairs switching to an allocentric navigational strategy. *Front Aging Neurosci* 4, 29, doi:10.3389/fnagi.2012.00029 (2012).
- Latini-Corazzini, L. et al. Route and survey processing of topographical memory during navigation. *Psychol Res* 74, 545-559, doi:10.1007/s00426-010-0276-5 (2010).
- Shikauchi, Y. & Ishii, S. Decoding the view expectation during learned maze navigation from human fronto-parietal network. *Scientific Reports* 5, 17648, doi:10.1038/srep17648 (2015).
- Vandenberg, S. G. & Kuse, A. R. Mental rotations, a group test of three-dimensional spatial visualization. *Percept Mot Skills* 47, 599-604, doi:10.2466/pms.1978.47.2.599 (1978).
- Viard, A., Doeller, C. F., Hartley, T., Bird, C. M. & Burgess, N. Anterior hippocampus and goal-directed spatial decision making. *J Neurosci* 31, 4613-4621, doi:10.1523/jneurosci.4640-10.2011 (2011).
- Voermans, N. C. et al. Interaction between the human hippocampus and the caudate nucleus during route recognition. *Neuron* 43, 427-435, doi:10.1016/j.neuron.2004.07.009 (2004).
- Wiener, J. M., de Condappa, O., Harris, M. A. & Wolbers, T. Maladaptive bias for extrahippocampal navigation strategies in aging humans. *J Neurosci* 33, 6012-6017, doi:10.1523/jneurosci.0717-12.2013 (2013).
- Wunderlich, K., Smittenaar, P. & Dolan, R. J. Dopamine enhances model-based over model-free choice behavior. *Neuron* 75, 418-424, doi:10.1016/j.neuron.2012.03.042 (2012).





# Of Reward, Age, and Space: Reinforcement Learning Algorithms Explain Age-related Changes in Spatial Navigation

**Dian Anggraini**<sup>1,3\*</sup>, Stefan Glasauer<sup>2,3</sup>, Virginia Flanagin<sup>2,3</sup>, Ot Prat<sup>3</sup>, Klaus Wunderlich<sup>1,3</sup>

1: Department of Psychology, Ludwig-Maximilians University of Munich

2: Center for Sensorimotor Research, Klinikum der Uni Muenchen

3: Graduate School of Systemic Neurosciences, Ludwig-Maximilians University of Munich

## **Authors' Contribution:**

Conceptualization: D.A., S.G., V.F., and K.W.; Methodology: D.A, S.G., V.F., and K.W.;  
Software: D.A. and K.W.; Validation: D.A, S.G., and K.W.; Resources: S.G.;  
Investigation: D.A. and O.P.; Formal Analysis: D.A., O.P., and K.W.; Writing – Original  
Draft: D.A. and O.P.; Writing – Review and Editing: D.A., S.G., V.F., and K.W.;  
Visualization: D.A. and K.W.; Supervision: S.G. and K.W; Project Administration: D.A.;  
Funding Acquisition: S.G. and K.W.

\*D.A is the sole first author. The manuscript is currently under final preparation for journal submission.

### 3.1 Abstract

The ability to find one's way in a complex environment is crucial to everyday functioning. This navigational ability relies on the integrity of several cognitive functions and different strategies, route and map-based navigation, that individuals may adopt while navigating in the environment. As the integrity of these cognitive functions often decline with age, navigational abilities show marked changes in both normal aging and dementia. Combining a wayfinding task in a virtual reality (VR) environment and modeling technique based on reinforcement learning (RL) algorithms, we investigated the effects of cognitive aging on the selection and adoption of navigation strategies in human. The older participants performed the wayfinding task while undergoing functional Magnetic Resonance Imaging (fMRI), and the younger participants performed the same task outside the MRI machine. Compared with younger participants, older participants traversed a longer distance. They also exhibited a higher tendency to repeat previously established routes to locate the target objects. Despite these differences, the traversed paths in both groups could be well explained by the model-free and model-based RL algorithms. Furthermore, neuroimaging results from the older participants show that BOLD signal in the ventromedial prefrontal cortex (vmPFC) pertained to model-free value signals. This result provide evidence on the utility of the RL algorithms to explain how the aging brain computationally prefer to rely more on the route-based navigation.

## 3.2 Introduction

Successful navigation requires not only spatial knowledge but also the adoption of appropriate strategies. Two separate strategies, namely the route-based and map-based navigation, can be used to navigate in the environment (Packard et al., 1989; Packard and McGaugh, 1996; Iaria et al., 2003; Voermans et al., 2004). The former relies on route knowledge and egocentric spatial reference frame. It critically involves learning a series of stimulus-response associations, such as patterns of right and left turn from a specific starting position (White and McDonald, 2002). The map-based navigation, in contrast, form an internal representation of relationships between different objects. This strategy, which relies on survey knowledge and allocentric spatial reference frame, is characterized by flexibility. In other words, it allows one to derive a novel path from a given starting position (Bohbot et al., 1998; Maguire et al., 1998; Naveh-Benjamin, 2000; Davachi and Wagner, 2002). Adopting either the route-based or map-based navigation involves complex cognitive faculties that may deteriorate in both normal aging and dementia.

Furthermore, brain areas that play a key role in map-based navigation, such as the hippocampus and medial temporal lobe (MTL), exhibit more extensive degradation in aging than other brain regions (Driscoll et al., 2003; Du et al., 2003). These regions also showed reduced functional activation during navigation with age (Moffat et al., 2006). Some aspects of navigation, such as repeating previously established route (i.e. route-based navigation) have been reported to be less susceptible to aging (Jansen et al., 2010). Consequently, older adults may need to rely more on the route-based navigation. It is therefore not surprising that many studies have reported the difficulties in adapting map-based navigation among older adults (Moffat and Resnick, 2002; Moffat et al., 2006; Iaria et al., 2009; Wiener et al., 2011; Harris et al., 2012; Harris and Wolbers, 2012; Wiener et al., 2013). However, the specific computational mechanism, accounting for route-based navigation preference in older adults remains a matter of debate.

In this study, we combined modeling techniques from value-based decision making (model-based and model-free reinforcement learning) and wayfinding task in a virtual reality (VR) environment to (1) test the hypothesis that cognitive aging results in a shift away from map-based navigation, (2) identify the computational mechanism accounting for this shift, and

(3) identify neural correlates pertaining to the preference of route-based navigation among older adults.

The core of our wayfinding task was a 4 by 4 grid world. Each room within this arena was distinctively furnished to allow participants to distinguish individual room (Figure 3.1. A and B). We conducted our experiment on two groups of participants: older adults between the age of 60 to 75, and younger adults between the ages of 23 to 35. The older participants performed the same task while undergoing functional Magnetic Resonance Imaging (fMRI). The younger participants performed the wayfinding task on a 24-inch monitor outside the MRI scanner. Since backtracking (leaving the room from the same door as entering it) was not allowed, rooms in the middle of the grid-world had three doors to choose from, rooms along the outside wall had two doors, and corner rooms had only one.

Each participant performed one training phase and three test phases of the wayfinding task. Participants navigated by freely choosing one of the available doors. During the training phase, participants were asked to follow the instructions appearing on the screen. Instructions indicated which door to take in each decision point, and thus ensured that participants visited all the rooms in the arena. The paths followed during this training phase were the same for all participants. These paths did not correspond to the paths the participants should take to retrieve the rewards, and the rewards were hidden from participants' plain sight. The main purpose of this phase was to allow all participants to get used to navigating with the button press and to explore the virtual environment. The test phase includes (1) encoding, (2) retrieval, and (3) search. During the encoding phase, participants always started in the same starting position and were asked to collect three rewards in a specific order over eight trials. During the retrieval phase, participants were instructed to collect one specific reward at a time from the same starting position as in the encoding phase. Rewards were positioned at the same location as in encoding phase and selected in random order. Participants completed 12 trials in the retrieval phase. During the search phase, participants had to collect a specific randomly chosen reward, each time starting from a different starting position. Participants completed 15 trials in the search phase (Figure 3.1. A and B).

In our task, we expected that younger participants would use a mixture of route and map-based navigation across trials, whereas the older participants would exhibit strong



propensity towards route-based navigation. We also expected that the traversed paths in both groups of participants could be well explained by the model-free and model-based RL algorithms. Significant correlations between navigation indices and key parameters from the RL models would then confirm that these algorithms account for individuals' variability in strategy adoption during navigation. Moreover, we expected that correlates between BOLD activity and RL models' key internal variables, such as value signals may provide evidence that these algorithms explain how the aging brain computationally prefer to rely more on route-based navigation.

### 3.3 Results

#### 3.3.1 Spatial Navigation Strategies in Two Groups of Participants

As a measure of spatial ability, participants performed paper-based Mental Rotation Task (MRT, details in the Supplementary Method). This test measures a specific visuo-spatial ability, which involves the process of imagining how a two- or three-dimensional object would look if it is rotated away from its original upright position. Using an independent sample t-test to compare the two groups of participants, we found a significant difference ( $t_{(58)} = 2.008, P < 5 \times 10^{-13}$ ) between younger ( $M = 25.97, SD = 9.67, N = 42$ ) and older ( $M = 10.54, SD = 2.36, N = 22$ ) participants with the younger participants performed better than the older cohorts (Figure 3.3. A and Supplementary Table 3.1, 3.2, and 3.3). This result is consistent with previous studies reporting a large age difference in mental rotation performance (Kaltner and Jansen, 2016). The older participants however, performed better during the post navigation test (Younger Participants:  $M = 10.93, SD = 2.88, N = 42$ ; Older Participants:  $M = 12.91, SD = 2.88, N = 22; t_{(58)} = 2.001, P < 0.002$ ; Figure 3.3. B).

During the encoding phase, most participants started by exploring the environment. This was often seen by moving from one side of the arena to the opposite until all rewards were collected. On the one hand, half of younger participants either found the shortest paths to go from one reward to the next, or established a certain route to go from one reward to the other. On the other hand, all older participants established a certain route to go from one reward to the other. This trend was also observed during the retrieval and search phase.

As a first crude measure to quantify strategy adoption in our subjects, we calculated two navigation indices based on: (1) the number of trials in which the shortest path was used (adapted from (Marchette et al., 2014; Marchette and Vass, 2015) and referred to as  $I_{PATH}$  in this work), (2) the number of repeated route trials (referred as  $I_{ROUTE}$ ). Each navigation index was first calculated for each subject individually on every phase of the wayfinding task. While we could distinguish whether subjects used novel shortest paths to retrieve rewards or simply followed the route during the search phase, this distinction was sometimes not so clear cut for the retrieval phase. This is because subjects had to find each reward from the same starting position across several trials. Therefore, to get an overview on how every subject performed throughout the entire experiment, every navigation index was then averaged across three phases of the wayfinding task. For  $I_{PATH}$ , a value of 1 indicates that a

subject would have used the shortest path on every trial. In other words, the participant was primarily displaying a map-based strategy. This is because, by design, only subjects who had a good spatial representation of the environment could have reached the rewards using the shortest paths. In contrast, participants whose scores were close to 0 predominantly used suboptimal long routes, suggesting that they lacked a map-like representation of the environment. In contrast, for  $I_{ROUTE}$ , a score of 1 indicates that a participant repeated previously established route, i.e. showed a tendency towards route-based strategy.

Looking at the distribution of  $I_{PATH}$  and  $I_{ROUTE}$  across different phases of the wayfinding task, the usage of optimal paths (as represented by higher  $I_{PATH}$ ) improved over experimental phase for younger participants. In contrast, the older cohorts showed little change in strategy use over the three phases of the wayfinding task.

To quantify strategy preference and changes thereof, we ran separate ANOVAs for each index ( $I_{PATH}$  and  $I_{ROUTE}$ ) with age (young vs. old) as between group factor and phase (encoding, retrieval, and search) as within-subjects factor. Main effect of age revealed that young participants relied more strongly on the map-based navigation than older participants ( $F_{(1,62)} = 84.99, P < 2 \times 10^{-16}$ ). The main effect of phase ( $F_{(1,62)} = 48.04, P < 2 \times 10^{-16}$ ) was driven by the strong increase in the use of optimal paths by the younger participants, which is reflected in the significant age  $\times$  phase interaction ( $F_{(2,62)} = 4.62, P < 0.02$ ). As for  $I_{ROUTE}$ , we also found main effect of age. The older group, compared with younger participants, relied more strongly on the route-based navigation ( $F_{(1,62)} = 73.22, P < 4.26 \times 10^{-15}$ ). There were also significant main effect of phase ( $F_{(1,62)} = 117.69, P < 2 \times 10^{-16}$ ) and phase  $\times$  age interaction ( $F_{(2,62)} = 17.39, P < 7.5 \times 10^{-8}$ ).

To get an overview on how the two groups differed in performance throughout the entire experiment, we compared the average of  $I_{PATH}$  and  $I_{ROUTE}$  (across three phases of the wayfinding task) between the two groups (Figure 3.3. C and D). This revealed that, as a group, younger participants showed a higher tendency toward map-based navigation, i.e., higher  $I_{PATH}$  score ( $t_{(58)} =, P < 1.34 \times 10^{-11}$ ). In contrast, older participants showed an overall preference for route-based strategy, i.e., higher  $I_{ROUTE}$  ( $t_{(58)} =, P < 6.92 \times 10^{-11}$ ). Such systematic preference for one strategy strongly suggests that the suboptimal longer routes taken by the older participants was not unspecific. Rather, it showed that they relied more on route-based navigation.

Furthermore, it has been traditionally accepted that males have an overall better performance than females in various tasks that involve spatial skills (Keeley et al., 2013; Sneider et al., 2015). Therefore, there is a potential of gender effect in our findings. To rule out this effect, we also compared strategy adoption between our male and female cohorts among our younger adult participants. We found no significant differences (details in the Supplementary Results and Discussion) were found among the two groups in either the optimal phase ( $I_{PATH}$  Female:  $M = 0.39$ ,  $SD = 0.19$ ; Male:  $M = 0.49$ ,  $SD = 0.22$ ; two-tailed t-test  $P = 0.133$ ), or the repeated runs ratio ( $I_{ROUTE}$  Female:  $M = 0.24$ ,  $SD = 0.10$ ; Male:  $M = 0.26$ ,  $SD = 0.12$ ; two-tailed t-test  $P = 0.529$ ).

### 3.3.2 RL Model Fits to Navigation Data

To computationally assess subjects' navigation strategy at every decision point, we modelled subjects' choice behaviour by fitting three different reinforcement learning (RL) algorithms: model-free, model-based, and a hybrid model. The hybrid model is formulated as a weighted combination of the model-free and the model-based algorithms. These algorithms exemplify two strategies in value-based decision making: the model-based choice that creates a cognitive representation of the entire environment, and the model-free choice that simply increases action values along the taken paths that previously led to rewards (Figure 3.2). The hybrid model assumes that participants would employ both algorithms at a relative degree, represented by a fitted parameter weight ( $\omega$ ) (Glascher et al., 2010; Daw et al., 2011). We performed model fitting for each participant individually and assessed the relative goodness of fit in every phase (for details see Methods, Supplementary Methods, and Supplementary Table 3.4 and 3.5).

In the younger participants, the fitted  $\omega$  (fitted individually for every phase of the wayfinding task, then averaged over the three different phases) from the hybrid model ( $M = 0.42$ ,  $SD = 0.26$ ,  $N = 42$ ) significantly correlated with spatial cognitive ability (as assessed by MRT,  $r_{(40)} = 0.46$ , two-tailed t-test  $P < 0.0025$ ) and  $I_{PATH}$  ( $M = 0.44$ ,  $SD = 0.21$ ,  $r_{(40)} = 0.66$ , two-tailed t-test  $P < 1.98 \times 10^{-6}$ ). We did not find significant correlation between fitted  $\omega$  (averaged over three different phases) in the hybrid model ( $M = 0.25$ ,  $SD = 0.15$ ,  $N = 22$ ) and the MRT score ( $r_{(20)} = -0.015$ , two-tailed t-test  $P < 0.95$ ) in the older cohort. Nevertheless, in older cohort, we found significant correlation between  $\omega$  and  $I_{PATH}$  ( $M =$

0.14,  $SD = 0.08$ ,  $r_{(20)} = 0.49$ , two-tailed t-test  $P < 0.02$ ). These results demonstrate that, across three phases of the wayfinding task, those who often took the shortest paths also showed a tendency towards the model-based choice (Figure 3.3. E and F).

### 3.3.3 BOLD Signals Pertaining to Value of Chosen Action

We then investigated neural responses pertaining to choice valuations in every room. A key internal variable of the RL algorithms is the value of the chosen action, a signal that has been reliably detected in BOLD fluctuations over a large number of studies (Wunderlich et al., 2009; Glascher et al., 2012; Kahnt et al., 2014). For each decision point, we calculated the values of the chosen action (i.e. go to the left, right, or straight ahead) based on model-based and model-free RL algorithms. We took these computed values along the traversed paths and used them as parametric modulators to analyse the fMRI data.

We analysed 3 different phases of the wayfinding task in one GLM instead of analysing each phase separately. The reason for this is that in even though there was a main effect of age  $\times$  phase in our wayfinding task, this main effect was mainly driven by the younger participants who exhibit a tendency toward map-based navigation during the retrieval and search phase. Since we did not find this tendency in the older participants, and to improve statistical power when we analysed the fMRI data, we aggregate our GLM analysis into one.

We found that BOLD activity correlated significantly with the model-free value signals along bilateral medial prefrontal cortex [mPFC, peak in the right hemisphere (superior medial gyrus):  $x = 6, y = 42, z = 40, T = 10.35, k = 3559$ ; peak in the left hemisphere (medial frontal gyrus):  $x = -6, y = 28, z = -10, T = 6.11, k = 132$ ] extending to right anterior cingulate cortex [ACC, peak:  $x = 4, y = 34, z = 20, T = 8.78, k = 3559$ ]. Other clusters included retrosplenial complex [ $x = 4, y = -24, z = 38, T = 9.23, k = 4729$ ] and precuneus [ $x = -6, y = -58, z = 28, T = 8.37, k = 4729$ ]. We did not find correlates of model-based signals that survived correction for multiple comparisons. See Figure 3.6 for the activated areas. For correction of multiple comparisons, we set our significance threshold at  $P < 0.05$  whole-brain FWE corrected for multiple comparison at cluster level.

### 3.4 Discussion

Previous studies have reported age-related changes in the selection and subsequent use of navigation strategies as well as neural computation and representation of space (Schuck et al., 2015; Lester et al., 2017). On the one hand, younger adults are adept at choosing an appropriate strategy based on their wayfinding needs. On the other hand, older adults often employed route-based navigation irrespective of task demands (Bohbot et al., 2012; Rodgers et al., 2012). This change in strategy preference across the adult lifespan has primarily been attributed to aged-related decline in hippocampal function that affects one's ability to acquire and utilize allocentric spatial reference frame and survey knowledge (Moffat and Resnick, 2002; Rosenzweig and Barnes, 2003; Raz et al., 2005; Moffat et al., 2006; Moffat et al., 2007; Moffat, 2009). In contrast, route-based navigation, which relies on egocentric reference frame, is relatively unaffected by age (Raz et al., 2004; Raz et al., 2005; Raz et al., 2010). In line with these findings, we found that younger participants used a mixture of route and map-based navigation across trials, whereas the older cohorts exhibited substantial tendency towards route-based navigation. Moreover, the traversed paths in both groups could be well explained by the model-free and model-based RL algorithms. Importantly, fMRI results from the older participants show correlated between BOLD signal in the mPFC and the model-free value signals from the RL algorithm.

In this study, we combined VR-based wayfinding paradigm, RL algorithms, and fMRI to investigate the effects of cognitive aging on the adoption of navigation strategies. Our approach differs from others used in aging research (Moffat and Resnick, 2002; Moffat et al., 2006; Moffat et al., 2007; Moffat, 2009; Wiener et al., 2009; Wiener et al., 2011; Wiener et al., 2012; Wiener et al., 2013), in a number of important aspects. First, participants were not confined to one strategy over the other. This means navigational strategy was measured in conditions that allowed either one of them or combination of both to prevail. To a certain degree, this is more comparable to the real-world situation. Furthermore, our encoding phase emphasized the formation of a route between a single starting position and three reward locations. Nevertheless, this phase still allowed the formation of a map-like representation, because participants could freely move within the environment. The retrieval phase then allowed further memory formation of how each of the three rewards can be reached from the starting position. During the search phase, different starting positions allowed participants to plan the shortest path using a map-like representation.

Finally, we also calculated behavioral measures. These behavioral measures were calculated (1) based on whether subjects took the shortest path, and (2) by modelling subjects' choice by choice progression using model-free, model-based, and hybrid model algorithms of RL. Significant correlation between these two behavioral measures show that choice behaviors of participants who could find the shortest path were better explained by the model-based algorithm. This suggests that, in both groups, the hallmarks of navigation strategies are seen at the population level and within many individuals. These hallmarks of navigation strategies are well captured by the reinforcement learning models. It is important to note that there might be between-subject variability in their deployment.

During the retrieval and search phase of the wayfinding task, participants navigated to find one specific reward per trial. To do so optimally, they needed to use map-based navigation. Looking at the distribution of the navigation indices, more than half of our younger participants showed clear evidence for integrating map-based navigation. This trend was also supported by a significant correlation of  $I_{PATH}$  and fitted parameter  $\omega$ . In contrast, only 3 out of 22 older participants integrated map-based navigation while searching for rewards during the retrieval and search phase. A possible explanation for the discrepancy between the two groups is age-related impairment in both formation and the use of a cognitive map. Iaria et al. (2009) reported that older adults required more time to form cognitive map. In addition, compared to younger adults, older adults took more time and made more errors when they used the cognitive map to solve the navigation task (Iaria et al., 2009). This finding is supported by more recent studies that point towards older adults' inability to utilize viewpoint-independent survey knowledge (Taillade et al., 2015; Lester et al., 2017).

Furthermore, empirical research in spatial navigation (Harris et al., 2012; Harris and Wolbers, 2014) and other cognitive domains (Kray and Lindenberger, 2000; Terry and Sliwinski, 2012; Butler and Weywadt, 2013) has demonstrated that older adults are more vulnerable to switching costs. Given these findings, one could hence argue that the discrepancy between older and younger participants could also reflect an inability to switch from a firmly established route-based navigation to a map-based navigation. In our study, we observed that older participants, when they started from a new starting position, often retraced a path that might lead them to a familiar location. From this location, they then followed the established path that lead them to the target reward. These results make the



strategy switching account unlikely and point instead to an age-related deficit in building a cognitive map in novel environments.

Going beyond the behavioral results, it is noteworthy that most attention in research and one age differences in human navigation has focused on the role of hippocampus and associated structures. This focus comes as no surprise in light of the prominent stage of the MTL in models of human spatial and episodic memory. It is important to note, however, findings from numerous neuroimaging studies reveals that navigation elicits activation in widespread cortical and subcortical regions beyond the hippocampus. In line with these findings, we found correlate of model-free values in the prefrontal region, i.e. mPFC and ACC, and the retrosplenial complex (RSC). This result has two important implications. First, it supports observation from both anatomical (Rajah et al., 2011) and functional neuroimaging studies that suggest that successful navigation in humans requires substantial contributions from the prefrontal circuit and associated cognitive systems (Chersi and Burgess, 2015; Dumont and Taube, 2015). Study by Moffat et al. (2007), for example, found that larger volumes of the prefrontal cortex grey and white matter were positively correlated with navigational performance in older adults. Furthermore, studies in nonhuman species confirm important contributions from the (medial) prefrontal regions in solving the Morris Water Maze (MWZ). Using the route and map-based version of the task, respective roles of the frontal cortex and hippocampal system have been delineated. Some studies even reported double dissociations in a way that hippocampal lesions impair map-based navigation and frontal cortex lesions impair route-based navigation (de Bruin et al., 1997; de Bruin et al., 2001).

The second important implication is that our fMRI results argue against a strictly hippocampal/memory explanation of age related changes. This notion is supported by both behavioral and fMRI findings. Several behavioral observations suggest that age-related deficit in wayfinding task cannot be fully accounted by models of performance focusing solely on spatial memory. In particular, some studies reported that healthy older adults performed worse than their younger counterparts even during the early trials of the wayfinding task when memory for the target location was not yet crucial (Moffat and Resnick, 2002; Driscoll et al., 2005). Likewise, in our study, older participants performed significantly better in the post navigation task. In other words, they have better memory of objects that they encountered during the task. Lastly, we also found that older participants

performed significantly worse in the mental rotation task. This could be an indication that decline in basic spatial ability contributes to age-related changes in navigation. Taken together, we could speculate that the discrepancy observed among older and younger adults may be partially attributed to impaired executive and strategic functions.

Equally important, several studies reported reduced or completely absent hippocampal activation among the older adults (Meulenbroek et al., 2004; Moffat et al., 2006; Antonova et al., 2009). In particular, when compared to younger adults, older participants showed reduced activation in the hippocampus and parahippocampal gyrus and medial parietal. In contrast, BOLD activity in the medial frontal lobe is relatively higher. In line with these findings, we also did not find correlates between RL's key internal variables and the BOLD signal in the MTL. Therefore, it would be critical to understand how different cortical systems work in concert with the hippocampal system. This understanding may further delineate how changes in the neural systems may manifest in the specific components and sub-processes responsible for age-related discrepancy in strategy adoption during spatial navigation.

In summary, the results of the present study replicate extant findings of age-related differences in spatial navigation and contribute new evidence toward understanding of their neural signatures. Of particular importance, these results indicate that successful navigation does not depend solely on the medial temporal lobe. Instead, it is associated with multiple cortical and subcortical structures and draws on task-specific spatial memory resources.

## **3.5 Methods**

### **3.5.1 Participants**

Sixty-four participants [42 younger (22 females; mean age: 26.3 years); 22 older (10 females; mean age: 62.6 years)] took part in the experiment. The Mini Mental State Examination (MMSE) was administered to all participants to screen for mild cognitive impairment (MCI). Three additional older participants and two additional younger participants were not included in the analysis. All participants had normal or corrected-to-normal vision and no history of either neurological/psychiatric illness or any other contraindications to the MRI environment.

We had a larger cohort of younger participants because in addition to comparing strategy adoption between younger and older adults, we also examined the possibility of gender effects in strategy adoption in the group of our younger adult participants. Traditionally, it has been accepted that males have an overall better performance than females in tasks involving spatial skill (Keeley et al., 2013; Sneider et al., 2015). However, these results could be attributed to gender stratification, that is, the difference in opportunities that women have in different cultures (details on possible gender effects in strategy adoption is provided in the Supplementary Results and Discussion).

### **3.5.2 Procedure**

Participants provided information on their age, gender, and computer experience prior to the experiment. The younger participants performed the wayfinding task on a standard desktop computer with a 24-inch widescreen monitor. The older participants performed the wayfinding task while undergoing a functional Magnetic Resonance Imaging. The wayfinding task was designed to assess the route-based and map-based strategy adoption (details in Supplementary Methods). In addition to the wayfinding task, participants also performed Mental Rotation Task (MRT) and post-navigation test (details on both tasks in Supplementary Method).

The study was approved by the ethics committee of the Medical Faculty of the Ludwig-Maximilian-University Munich. Participants were made fully aware of the details of the

experimental procedures and gave written informed consent. Younger participants were paid a compensation of 20€ and older participants were paid a compensation of 100€ to partake in the study, which took 70 to 90 minutes to complete.

### 3.5.3 Navigation Indices

To quantify strategy adoption during the wayfinding task, we calculated three navigation indices. For each navigation index, we calculated the index separately for each phase of the wayfinding task (adapted from (Marchette et al., 2011)). We first measured (1) the number of trials when participants used the shortest path to find the rewards, and (2) number of trials where participants repeated learned routes.

#### *I<sub>PATH</sub>*

We then calculated the proportion of trials with the shortest path to the total number of trials in a respective run

$$I_{PATH} = \frac{n_{shortest\_trial}}{n_{total\_trial}} \quad [3.1]$$

$n_{shortest\_trial}$  is the number of trials when participants used the shortest path to find the rewards and  $n_{total\_trial}$  is the number of trials in a given phase. A score of 1 on this index suggests that a subject used the shortest path on every trial, i.e., the subject was primarily displaying a map-based strategy. In contrast, a score of 0 indicates that a subject always used a suboptimal longer path.

#### *I<sub>ROUTE</sub>*

To analyze whether participants simply followed previously established route to reach reward, we also calculated the proportion of trials with the shortest paths and trials where participants simply repeated previously learned routes. Note that for the encoding phase, this index might be less meaningful because participants were still learning the environment and their task was to establish a certain route. We defined this index as

$$I_{ROUTE} = \frac{n_{route\_trial}}{n_{route\_trial} + n_{shortest\_trial}} \quad [3.2]$$

$n_{route\_trial}$  is the number of trials when participants repeated learned routes.  $I_{PATH}$ ,  $I_{STEPS}$ , and  $I_{ROUTE}$  were first calculated for each phase separately. To get a general idea on how participants navigate throughout the entire experiment, we averaged each index over three phases of the wayfinding task.

### 3.5.4 Reinforcement Learning Model

We modelled the sequence of participants' choices ( $a_i$ ) by comparing them step by step to those predicted by different learning algorithms as having encountered the same state ( $s$ ), action ( $a$ ) and reward ( $r$ ). As we had 4 by 4 grid, the wayfinding task consisted of 16 states and in each state, participants could have up to three actions depending on which direction the subject was facing.

Since the wayfinding task consisted of three rewards, the goal for both model-free and model-based algorithms is to learn the state-action value function  $Q(s,a)$  mapping at each state-action pair to each reward. We assume no interference or generalization between the three rewards conditions, and thus each algorithm was subdivided into three independent task set, one for each reward.

#### ***Model-free reinforcement learning***

For model-free choice, we used the SARSA with eligibility traces (SARSA ( $\lambda$ )) to calculate model-free value or  $Q_{MF}$  (Sutton and Barto, 1998). This algorithm has three free parameters: learning-rate ( $\alpha$ ), inverse temperature ( $\beta$ ), and eligibility parameter ( $\lambda$ ). Each state-action pair is associated with a value  $Q_{MF}(s,a)$  all initially set to 0. The eligibility trace  $Z$ , set to 1 at the beginning of the trial and assumed not to be carried over from trial to trial, allows us to update each state-action pair along a subject's encountered trajectory. For every trial  $t$  in which the subject located the reward ( $r$ ), the state-action value is updated for each step  $i$  in that trial according to the following:

$$Q_{MF}(s_{i,t+1}, a_{i,t+1}) \leftarrow Q_{MF}(s_{i,t}, a_{i,t}) + \alpha \delta_{i,t} Z_i, \quad [3.3]$$

where

$$\delta_{i,t} \leftarrow R + Q_{MF}(s_{i+1,t}, a_{i+1,t}) - Q_{MF}(s_{i,t}, a_{i,t}), \quad [3.4]$$

and

$$Z(s_{i,t}, a_{i,t}) \leftarrow \lambda Z(s_{i,t}, a_{i,t}). \quad [3.5]$$

### ***Model-based reinforcement learning***

A model-based choice learns the configuration of the grid world and computes action values by searching across possible trajectories to locate the reward (Simon and Daw, 2011). Based on the grid-world configuration, we compute state-action values based on a planning process terminating at reward states. Specifically, for each action  $a$  in room  $s$ , we first initialized all  $Q_{MB}(s,a)$  to 0. Then, for all state-action pairs  $(s,a)$  and adjacent (next room) state-action pairs  $(s', a')$  we iteratively perform the following:

$$Q_{MB}(s, a) \leftarrow \begin{cases} R(s') & \text{if } R(s') \neq 0 \\ \max_{a' \in A} Q_{MB}(s', a') - (\gamma * \max_{a' \in A} Q_{MB}(s', a')) & \text{otherwise} \end{cases} \quad [3.6]$$

The algorithm has one fixed parameter  $\gamma$  that is set to 0.1. We took model-based values ( $Q_{MB}$ ) to be the values resulting after the algorithm converged (this occurred within 25 iterations). Note that computations of model-based value did not depend on the trial  $t$  or step  $i$  of the subject.

### ***Hybrid model***

In addition to model-free and model-based algorithm, we also considered a hybrid model (Glascher et al., 2010; Wunderlich et al., 2012) in which the model predicted values for the actions are calculated as a weighted linear combination of the values from model-free and model-based algorithms:

$$Q_{hybrid} = (1 - \omega)Q_{MF} + \omega Q_{MB} \quad [3.7]$$

The relative degree that the model-based algorithm contributed over the model-free is captured by the weight parameter ( $\omega$ ). We took this  $\omega$  as a free parameter, which was fitted individually for each subject but assumed to be constant throughout a single phase of the wayfinding task.

### ***Model fitting***

For each algorithm, we estimated a set of free parameters separately for each subject and for each phase of the wayfinding task by mean of hierarchical model fitting (Wunderlich et al., 2012). Further details on the model fitting and calculation of model evidence are provided in the Supplementary Methods. Estimation of the free parameters was implemented in Matlab using the Global Optimization Toolbox.

### **3.5.5 fMRI Data Analysis**

The following software were used for the fMRI data analysis:

SPM12 (<https://fil.ion.ucl.ac.uk>, RRID:SCR\_007037);

Xjview (<https://alivelearn.net/xjview/>, RRID:SCR\_008642);

SPM Anatomy Toolbox (<https://fz-juelich.de/>, RRID:SCR\_013273);

mricron (<https://nitrc.org/mricron>, RRID: SCR\_002403).

#### ***General Linear Model (GLM) for fMRI data analysis***

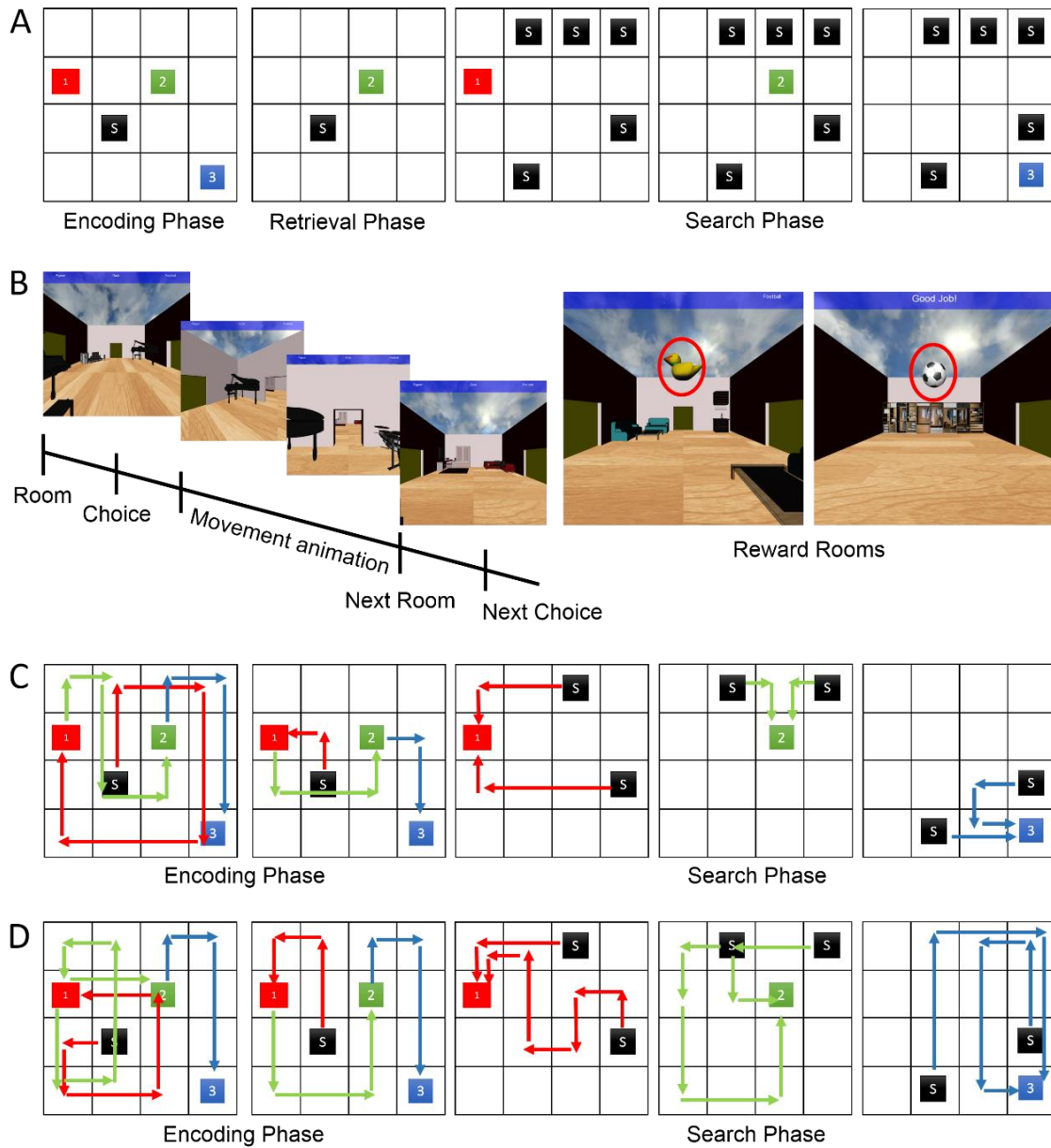
An event related analysis was applied on two levels using the general linear model approach as implemented in SPM12. Individual (random-effects) model parameters were used to generate regressors for the analysis of the fMRI data. The GLM included the following event regressors covering the time when participants saw (1) the instruction, (2) the room, (3) chose which direction they wanted to go (button press), (4) animation of movement, (5) seeing the reward, and (6) six seconds resting period in between trials. Our analysis focused on the times when subject entered each room and the button press to indicate where to go next. For our primary hypothesis and based on the results of the model-fitting, the decision time points were parametrically modulated by model-free values and (2) model-based values. Parametric regressors were not serially orthogonalized, thus allowing each regressor to account independently for the response at each voxel. Using this approach, we let the model-free and model-based value regressor directly compete for variance in the BOLD signal. In this approach, only variance that is exclusively explained by one or the other regressor is assigned to the regressor but not the variance that is shared by both.

All regressors were convolved with the canonical hemodynamic response function as provided by SPM12 and its temporal derivative. The six rigid-body motion parameters from head motion correction were also included in the model as regressors of no interest. At the first level, linear weighted contrasts were used to identify effects of interest, providing contrast images for group effects analysed at the second (random-effect) level.

*Second level analysis.* Calculated linear contrasts of parameter estimates, from the first level GLM analysis, for each regressor were then brought to the separate second level random-effects analysis, wherein one sample t-test provided effect for each regressor of interest. For correction of multiple comparisons, we set our significance threshold at  $P < 0.05$  whole-brain FWE corrected for multiple comparison at cluster level. The minimum spatial extent,  $k = 25$ , for the threshold was estimated based on the underlying voxel-wise  $P$  value with a cluster defining threshold of  $P = 0.001$ . Details on the fMRI preprocessing are provided in the Supplementary Methods.



### 3.6 Figures



**Figure 3.1. Wayfinding task and examples of participants' path**

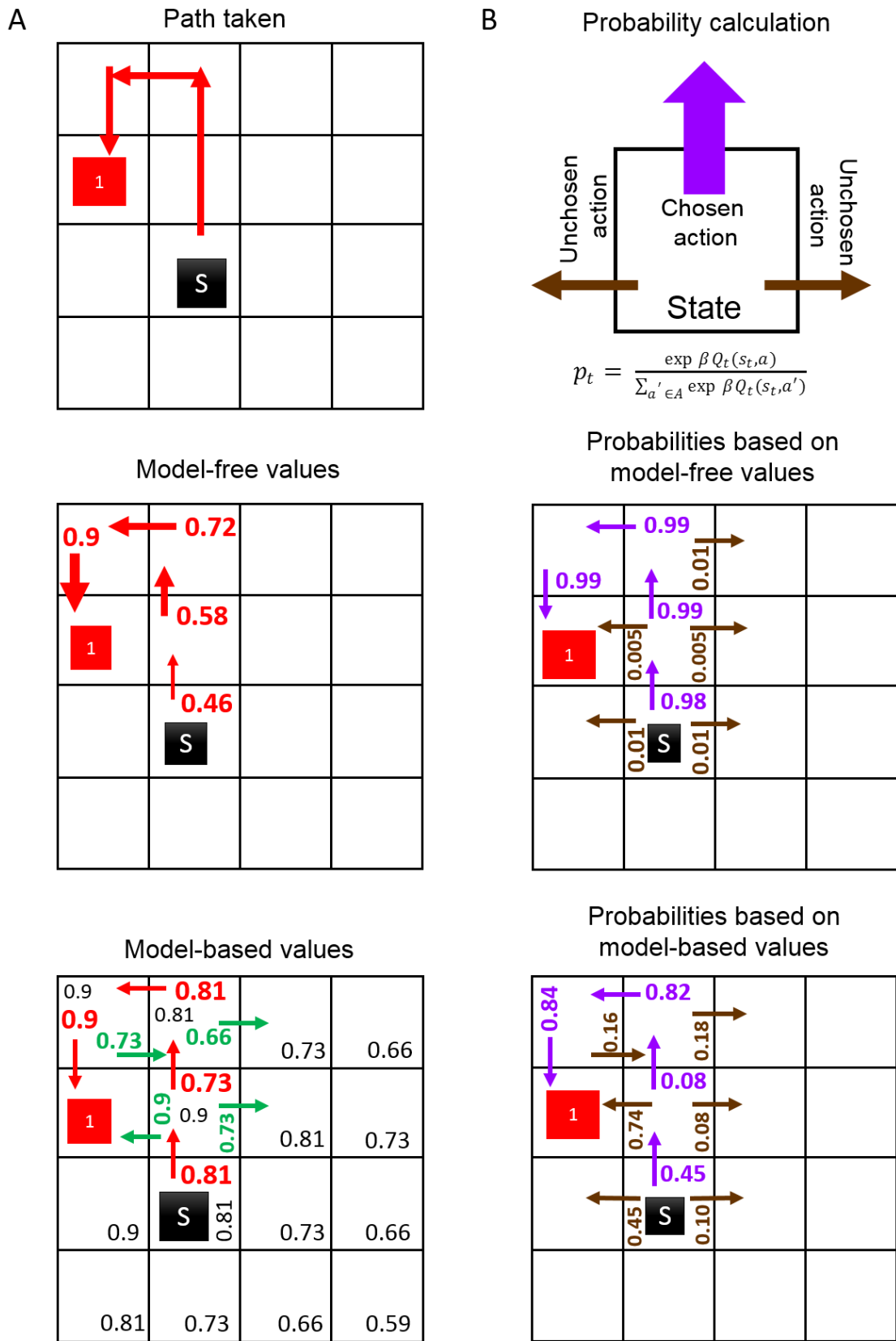
*A.* Layout of the grid-world. The Virtual Reality (VR) environment consisted of a 4 by 4 grid of rooms. Each square represents a room which contained distinct furniture and objects to distinguish individual rooms. Black square represents starting position, colored squares reward locations and the number represents the order in which they need to be found. The test phase of wayfinding task consisted of three phases: encoding, retrieval, and search phase. During the encoding phase, participants always started in the same starting position and were asked to collect three rewards in a specific order over eight trials. During the retrieval phase, participants were instructed to collect one specific reward at a time from the same starting position as in the encoding phase. During the search phase, subject had to locate one randomly chosen reward at each trial, each time starting from a different starting position.

*B.* Screenshots of the virtual reality environment. Each room is furnished with distinct furniture to allow participants to distinguish and recognize individual rooms. At each room (decision point) participants could choose up to three directions (corner rooms had either one or two directions to choose). After a choice was

made, an animation was leading to the room in the selected direction; this movement lasted 2.5 – 3 seconds jittered uniformly. The next room and, if applicable, the reward were presented.

*C.* Path from a representative younger participant who exhibited a tendency towards map-based strategy. During the encoding phase, the subject established by repetition a fixed route from one reward to the other. During the search phase, the participant managed to find the shortest path to reach the reward.

*D.* Path from a representative older participant who exhibited a tendency towards route-based strategy. During the encoding phase, the subject established by repetition a fixed route from one reward to the other. During the search phase, the subject still followed the established route to reach the reward.

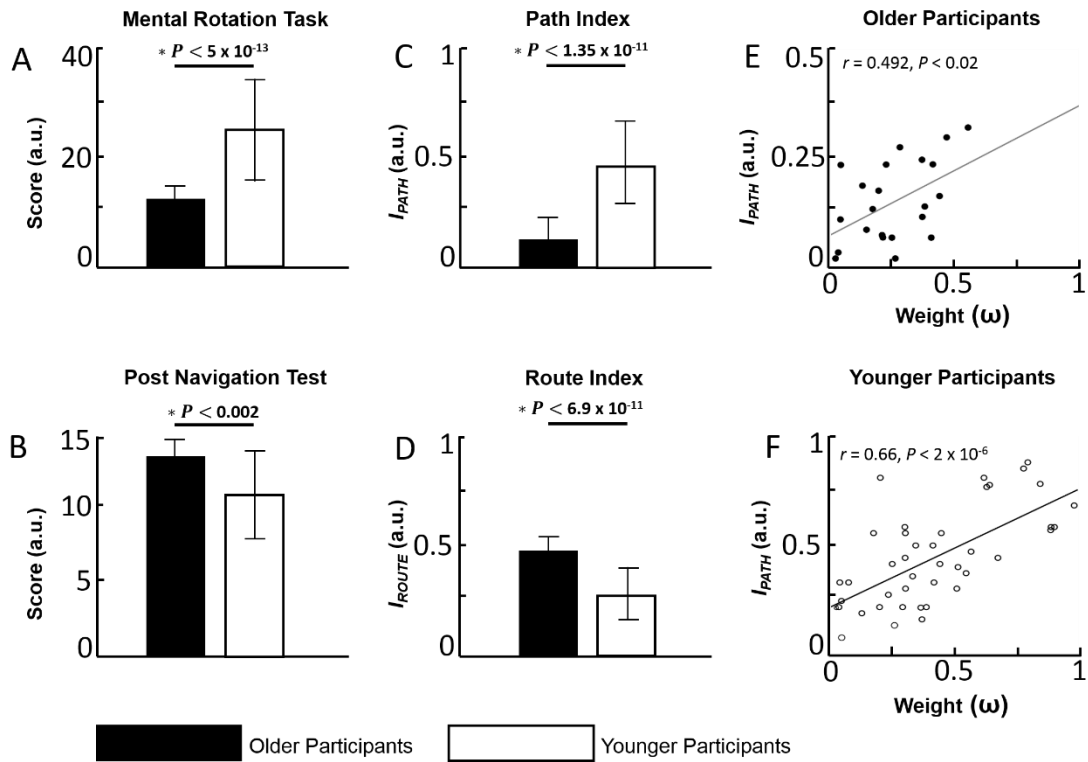


**Figure 3.2. Reinforcement learning models and model fitting**

A. Calculation of action values, showing how valuable it is to move along the route in a certain state, based on either model-free or model-based RL algorithm. The top panel displays the path the participant take to

reach the first reward. The middle panel displays model-free valuation based on the SARSA ( $\lambda$ ) algorithm. After reaching a reward this algorithm updates the values only along the traversed path. Lastly, the bottom panel displays how model-based algorithm updates values. The model-based RL algorithm, derived from dynamic programming, updates values not only along the taken path, but across the entire grid world in iterative manner. For this bottom panel, black numbers are state values, red numbers are model-based values of chosen action, and green numbers refer to model-based values of the unchosen actions.

*B.* Calculation of probability of taking certain actions in those state based on the action values. Purple numbers are probabilities of chosen action, brown numbers refer to probabilities of other unchosen actions. Note that not all probabilities for non-preferred actions are shown.



**Figure 3.3. Behavioral Results**

*A.* Performance on the MRT. Comparing MRT score between the older and younger participants, results indicate a significant difference ( $P < 0.0007$ ), with younger participants generally performing better than the older cohort.

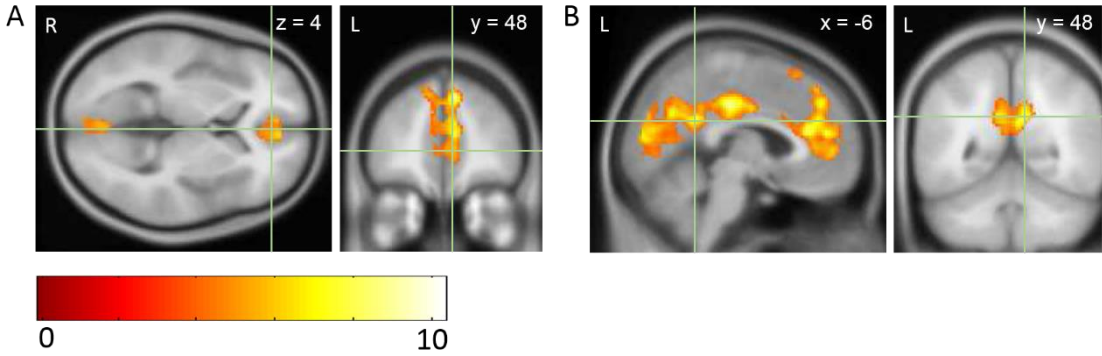
*B.* Performance on the Post Navigation Test. Comparing the score between the two groups of participants, results indicate a significant difference ( $P < 0.005$ ), with older participants generally performing better than the younger cohort.

*C.* Significant difference between the two group results ( $P < 1.3E-11$ ), older participants scoring (on average) lower than the younger participants.

*D.* Significant difference between the two group results ( $P < 7.1E-11$ ), older participants scoring (on average) higher than the younger participants.

*E.* Significant positive correlation between  $I_{PATH}$  and  $\omega$  parameters for the older group.

*F.* Significant positive correlation between  $I_{PATH}$  and  $\omega$  parameters for the younger group.



**Figure 3.4. BOLD activity pertaining to model-free value signals**  
A. Correlates of model-free valuations in medial prefrontal cortex (mPFC)  
B. Correlates of model-free valuations in retrosplenial complex (RSC)

### 3.7 References

- Antonova E, Parslow D, Brammer M, Dawson GR, Jackson SH, Morris RG (2009) Age-related neural activity during allocentric spatial memory. *Memory* (Hove, England) 17:125-143.
- Bohbot VD, Kalina M, Stepankova K, Spackova N, Petrides M, Nadel L (1998) Spatial memory deficits in patients with lesions to the right hippocampus and to the right parahippocampal cortex. *Neuropsychologia* 36:1217-1238.
- Bohbot VD, McKenzie S, Konishi K, Fouquet C, Kurdi V, Schachar R, Boivin M, Robaey P (2012) Virtual navigation strategies from childhood to senescence: evidence for changes across the life span. *Frontiers in aging neuroscience* 4:28.
- Butler KM, Weywadt C (2013) Age differences in voluntary task switching. *Psychology and aging* 28:1024-1031.
- Chersi F, Burgess N (2015) The Cognitive Architecture of Spatial Navigation: Hippocampal and Striatal Contributions. *Neuron* 88:64-77.
- Davachi L, Wagner AD (2002) Hippocampal contributions to episodic encoding: insights from relational and item-based learning. *Journal of neurophysiology* 88:982-990.
- de Bruin JP, Swinkels WA, de Brabander JM (1997) Response learning of rats in a Morris water maze: involvement of the medial prefrontal cortex. *Behavioural brain research* 85:47-55.
- de Bruin JP, Moita MP, de Brabander HM, Joosten RN (2001) Place and response learning of rats in a Morris water maze: differential effects of fimbria fornix and medial prefrontal cortex lesions. *Neurobiology of learning and memory* 75:164-178.
- Driscoll I, Hamilton DA, Yeo RA, Brooks WM, Sutherland RJ (2005) Virtual navigation in humans: the impact of age, sex, and hormones on place learning. *Hormones and behavior* 47:326-335.
- Driscoll I, Hamilton DA, Petropoulos H, Yeo RA, Brooks WM, Baumgartner RN, Sutherland RJ (2003) The aging hippocampus: cognitive, biochemical and structural findings. *Cerebral cortex* (New York, NY : 1991) 13:1344-1351.

- Du AT, Schuff N, Zhu XP, Jagust WJ, Miller BL, Reed BR, Kramer JH, Mungas D, Yaffe K, Chui HC, Weiner MW (2003) Atrophy rates of entorhinal cortex in AD and normal aging. *Neurology* 60:481-486.
- Dumont JR, Taube JS (2015) The neural correlates of navigation beyond the hippocampus. *Progress in brain research* 219:83-102.
- Glascher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66:585-595.
- Harris MA, Wolbers T (2012) Ageing effects on path integration and landmark navigation. *Hippocampus* 22:1770-1780.
- Harris MA, Wolbers T (2014) How age-related strategy switching deficits affect wayfinding in complex environments. *Neurobiology of aging* 35:1095-1102.
- Harris MA, Wiener JM, Wolbers T (2012) Aging specifically impairs switching to an allocentric navigational strategy. *Frontiers in aging neuroscience* 4:29.
- Iaria G, Palermo L, Committeri G, Barton JJ (2009) Age differences in the formation and use of cognitive maps. *Behavioural brain research* 196:187-191.
- Iaria G, Petrides M, Dagher A, Pike B, Bohbot VD (2003) Cognitive strategies dependent on the hippocampus and caudate nucleus in human navigation: variability and change with practice. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 23:5945-5952.
- Jansen P, Schmelter A, Heil M (2010) Spatial knowledge acquisition in younger and elderly adults: a study in a virtual environment. *Experimental psychology* 57:54-60.
- Kaltner S, Jansen P (2016) Developmental Changes in Mental Rotation: A Dissociation Between Object-Based and Egocentric Transformations. *Advances in cognitive psychology* 12:67-78.
- Kray J, Lindenberger U (2000) Adult age differences in task switching. *Psychology and aging* 15:126-147.
- Lester AW, Moffat SD, Wiener JM, Barnes CA, Wolbers T (2017) The Aging Navigational System. *Neuron* 95:1019-1035.



- Maguire EA, Burgess N, Donnett JG, Frackowiak RS, Frith CD, O'Keefe J (1998) Knowing where and getting there: a human navigation network. *Science (New York, NY)* 280:921-924.
- Marchette SA, Vass LK (2015) Outside Looking In: Landmark Generalization in the Human Navigational System. 35:14896-14908.
- Marchette SA, Bakker A, Shelton AL (2011) Cognitive mappers to creatures of habit: differential engagement of place and response learning mechanisms predicts human navigational behavior. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31:15264-15268.
- Marchette SA, Vass LK, Ryan J, Epstein RA (2014) Anchoring the neural compass: coding of local spatial reference frames in human medial parietal lobe. *Nat Neurosci* 17:1598-1606.
- Meulenbroek O, Petersson KM, Voermans N, Weber B, Fernandez G (2004) Age differences in neural correlates of route encoding and route recognition. *NeuroImage* 22:1503-1514.
- Moffat SD (2009) Aging and spatial navigation: what do we know and where do we go? *Neuropsychology review* 19:478-489.
- Moffat SD, Resnick SM (2002) Effects of age on virtual environment place navigation and allocentric cognitive mapping. *Behavioral neuroscience* 116:851-859.
- Moffat SD, Elkins W, Resnick SM (2006) Age differences in the neural systems supporting human allocentric spatial navigation. *Neurobiology of aging* 27:965-972.
- Moffat SD, Kennedy KM, Rodrigue KM, Raz N (2007) Extrahippocampal contributions to age differences in human spatial navigation. *Cerebral cortex (New York, NY : 1991)* 17:1274-1282.
- Naveh-Benjamin M (2000) Adult age differences in memory performance: tests of an associative deficit hypothesis. *Journal of experimental psychology Learning, memory, and cognition* 26:1170-1187.
- Packard MG, McGaugh JL (1996) Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiology of learning and memory* 65:65-72.

- Packard MG, Hirsh R, White NM (1989) Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: evidence for multiple memory systems. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 9:1465-1472.
- Rajah MN, Languay R, Grady CL (2011) Age-related changes in right middle frontal gyrus volume correlate with altered episodic retrieval activity. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31:17941-17954.
- Raz N, Ghisletta P, Rodrigue KM, Kennedy KM, Lindenberger U (2010) Trajectories of brain aging in middle-aged and older adults: regional and individual differences. *NeuroImage* 51:501-511.
- Raz N, Gunning-Dixon F, Head D, Rodrigue KM, Williamson A, Acker JD (2004) Aging, sexual dimorphism, and hemispheric asymmetry of the cerebral cortex: replicability of regional differences in volume. *Neurobiology of aging* 25:377-396.
- Raz N, Lindenberger U, Rodrigue KM, Kennedy KM, Head D, Williamson A, Dahle C, Gerstorff D, Acker JD (2005) Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. *Cerebral cortex (New York, NY : 1991)* 15:1676-1689.
- Rodgers MK, Sindone JA, 3rd, Moffat SD (2012) Effects of age on navigation strategy. *Neurobiology of aging* 33:202.e215-222.
- Rosenzweig ES, Barnes CA (2003) Impact of aging on hippocampal function: plasticity, network dynamics, and cognition. *Progress in neurobiology* 69:143-179.
- Schuck NW, Doeller CF, Polk TA, Lindenberger U, Li SC (2015) Human aging alters the neural computation and representation of space. *NeuroImage* 117:141-150.
- Simon DA, Daw ND (2011) Neural correlates of forward planning in a spatial decision task in humans. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31:5526-5539.
- Sutton RS, Barto AG (1998) Eligibility Traces. In: *Reinforcement Learning: An Introduction*. Cambridge: MIT Press.
- Taillade M, N'Kaoua B, Sauzeon H (2015) Age-Related Differences and Cognitive Correlates of Self-Reported and Direct Navigation Performance: The Effect of Real and Virtual Test Conditions Manipulation. *Frontiers in psychology* 6:2034.

- Terry CP, Sliwinski MJ (2012) Aging and random task switching: the role of endogenous versus exogenous task selection. *Experimental aging research* 38:87-109.
- Voermans NC, Petersson KM, Daudey L, Weber B, Van Spaendonck KP, Kremer HP, Fernandez G (2004) Interaction between the human hippocampus and the caudate nucleus during route recognition. *Neuron* 43:427-435.
- White NM, McDonald RJ (2002) Multiple parallel memory systems in the brain of the rat. *Neurobiology of learning and memory* 77:125-184.
- Wiener JM, Büchner SJ, Hölscher C (2009) Taxonomy of Human Wayfinding Tasks: A Knowledge-Based Approach. *Spatial Cognition and Computation* 9:13.
- Wiener JM, Berthoz A, Wolbers T (2011) Dissociable cognitive mechanisms underlying human path integration. *Experimental brain research* 208:61-71.
- Wiener JM, Kmecova H, de Condappa O (2012) Route repetition and route retracing: effects of cognitive aging. *Frontiers in aging neuroscience* 4:7.
- Wiener JM, de Condappa O, Harris MA, Wolbers T (2013) Maladaptive bias for extrahippocampal navigation strategies in aging humans. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33:6012-6017.
- Wunderlich K, Smittenaar P, Dolan RJ (2012) Dopamine enhances model-based over model-free choice behavior. *Neuron* 75:418-424.

## 3.8 Supplementary Information

### 3.8.1 Supplementary Results and Discussion

#### *Gender Differences in Spatial Learning and Navigational Abilities*

In addition to comparing strategy adoption between younger and older adults, we also examines the gender effects in strategy adoption in the group of our younger adult participants. Traditionally, it has been accepted that males have an overall better performance than females in tasks involving spatial skill (Keeley et al., 2013; Sneider et al., 2015). However, these results could be attributed to gender stratification, that is, the difference in opportunities that women have in different cultures. For instance, the spatial skills acquired in engineering careers (where males have traditionally been predominant) and videogames (also an activity usually associated with males) could account for most of these differences. Moreover, recent meta-analysis studies have shown how the divergence in spatial skills among genders is not as wide as it was usually assumed (Hyde, 2016) and can be reduced even further with proper training (Uttal et al., 2012).

Among the types of spatial skills, the biggest difference favouring males is found on performance in 3D mental rotation tasks (Hyde, 2016). Comparing male and female's MRT scores through a t-test, we found a significant effect of gender on MRT scores (Female:  $M = 21.80$ ,  $SD = 9.25$ ; Male:  $M = 29.77$ ,  $SD = 8.82$ ; two-tailed t-test  $P < 0.005$ ), with males generally performing better than females. This result allowed us to corroborate some previous finding in this field.

When comparing performance between the male and female younger adult participants during the wayfinding task, no significant differences were found among the two groups in either the optimal phase ( $I_{PATH}$  Female:  $M = 0.39$ ,  $SD = 0.19$ ; Male:  $M = 0.49$ ,  $SD = 0.22$ ; two-tailed t-test  $P = 0.133$ ), or the repeated runs ratio ( $I_{ROUTE}$  Female:  $M = 0.24$ ,  $SD = 0.10$ ; Male:  $M = 0.26$ ,  $SD = 0.12$ ; two-tailed t-test  $P = 0.529$ ).

Lastly, the post-navigation test aimed to estimate the importance participants gave to landmarks, and intended to be an indicator on how much they relied on landmark information during the wayfinding task. Comparing the test scores of the male and female younger adults participants, we found a significant difference between the two group results (Female:  $M = 12.05$ ,  $SD = 2.76$ ; Male:  $M = 9.91$ ,  $SD = 2.74$ ; two-tailed t-test  $P < 0.01$ ),

with females scoring higher than males. Higher scores on this test showed that females gave more attention to spatial landmarks, which are more closely related to route-based navigation as opposed to geometrical cues.

### **3.8.2 Supplementary Methods**

#### ***Wayfinding Task***

The Virtual Environment (VE) consisted of an arena divided into 16 square rooms (4-by-4 grid). The rooms were distributed in a grid-like manner, and each one of them was connected to the adjacent room. Every room was furnished in with unique set of furniture, making it different from all other rooms. While inside one of the room, no information from neighbouring room could be perceived.

The virtual environment was presented on a 24 inch computer screen (behavioral experiment) or 32 inch NNL screen (fMRI experiment). Participants navigated by pressing left, right, or forward button. There was no way of going directly backwards, and therefore, participants could not leave a room by the same door they just entered it.

The complete experiment consisted of one training phase and three test phases. During the training phase, participants were asked to follow the instructions appearing on the screen. Instructions indicated which door to take in each decision point, and allowed the participants to visit all the rooms in the arena. The paths followed during this training phase were the same for all participants. These paths did not correspond to the paths the participants should take to retrieve the rewards. The main purpose of this phase was to allow all participants to get used to navigating with the button press to suppress any variability caused by previous experience with this type of device. However, since the composition of the rooms remained unchanged, this training phase also gave all participants the opportunity to explore the virtual environment.

During the encoding phase, participants were instructed to find three rewards in the exact order. This was ensured by making each reward visible only if the previous reward had been already encountered. Once the participants had found all three rewards, they were brought back to the same starting position and had to perform the exact same task seven more times

(8 trials in total). During the retrieval phase, participants were only asked to find one reward at a time for 12 trials in total. Lastly, during the search phase, participants were, once again, instructed to find only the reward indicated on the screen. This time, however, the starting position for every new trial (15 trials in total) was different from the previous phase.

For the older adult participants, there was a practice session conducted three to five days before the fMRI scanning sessions. They practiced on a similar 4 by 4 grid world with different room furnishings but otherwise similar to the main task.

### ***Mental Rotation Task (MRT)***

The Mental Rotation Task (MRT) consisted on 24 target items, which were graphic representations of a three-dimensional object. Next to the target item, four additional testing figures were presented, and participants were required to identify two rotated images of the target item among those alternatives. After a training session where participants could get used to the task and practice with some additional examples, participants were given 10 minutes to complete as many items as possible. One point was awarded for each correct choice, resulting in a possible score ranging from 0 to 48.

### ***Post-navigation Test***

After the navigational task in the virtual environment, participants completed post-navigation test. The test was used to assess how many landmarks that the participants observed. The test had 15 questions, each of them consisting on 4 images representing different types of furniture. Among those four images, two of them represented furniture that were found in the virtual environment, while the other two were not present in the VE arena. Participants were asked to identify the two objects belonging to the VE arena, and a point was only awarded in case the subject successfully identified two correct options in a single test question. Therefore, scores ranged from 0 to 15.

### ***Model Fitting***

For each algorithm, we calculated a probability of choice based on the following softmax decision rule (Luce choice rule):

$$p_t = \frac{\exp \beta Q_t(s_t, a)}{\sum_{a' \in A} \exp \beta Q_t(s_t, a')} \quad [S3.1]$$

The parameter  $\beta$  is the inverse temperature representing the degree of stochasticity of participants' action selection.

We estimated a set of free parameters ( $\theta$ ) for each algorithm separately for each subject by mean of hierarchical model fitting (Wunderlich et al., 2012). First we applied logistic ( $\alpha$ ,  $\omega$ ,  $\lambda$ ) and exponential ( $\beta$ ) transformation before fitting parameters to transform bounded parameter into Gaussian distributed parameter values with population mean and standard deviation. In the equation below, the Greek alphabets represent the parameters we used in the model, while the Latin alphabets stand for parameters in the logistic transformations that range from  $-\infty$  to  $\infty$ .

$$\alpha = \frac{1}{1+\exp(-a)}, \quad \omega = \frac{1}{1+\exp(-w)}, \quad \lambda = \frac{1}{1+\exp(-l)}, \quad \beta = \exp(b) \quad [\text{S3.2}]$$

Next, we fitted these transformed parameters by maximizing the likelihood of all observed choices given the parameterized model:

$$L = P(c_i | \mu_a, \mu_l, \mu_b, \mu_w, \sigma_a, \sigma_l, \sigma_b, \sigma_w) = \int da_i dl_i db_i dw_i (c_i | a_i, l_i, b_i, w_i) (a_i | \mu_a, \sigma_a) P(l_i | \mu_l, \sigma_l) P(b_i | \mu_b, \sigma_b) P(w_i | \mu_w, \sigma_w) \quad [\text{S3.3}]$$

We later estimated mean and variance of the parameter distribution in the population based on our subject sample. As an example, for parameter  $\alpha$ :

$$\mu_a = \frac{1}{N} \sum_i^N a_i \quad \text{and} \quad \sigma_a = \sqrt{\frac{1}{N} \sum_i^N (a_i - \mu_a)^2} \quad [\text{S3.4}]$$

Following mean and variance estimation, we refitted single subject parameter values by minimizing over both the negative log likelihood of participants' choice given the parameters and the negative log likelihood for individual subject parameter values given the distribution of parameters in the population:

$$P(a_i, l_i, b_i, w_i | c_i, \mu_a, \mu_l, \mu_b, \mu_w, \sigma_a, \sigma_l, \sigma_b, \sigma_w) \propto (c_i | a_i, l_i, b_i, w_i) \times (a_i, l_i, b_i, w_i | \mu_a, \mu_l, \mu_b, \mu_w, \sigma_a, \sigma_l, \sigma_b, \sigma_w) \quad [\text{S3.5}]$$

### ***Model Evidence***

To approximate the model evidence, we computed the Bayesian Information criterion (BIC) as follows:

$$l(\hat{\theta}) + \frac{m}{2} \log n \quad [\text{S3.6}]$$

where  $l(\theta)$  is the negative log-likelihood of data at the maximum likelihood parameters  $\theta$ ;  $m$  is the number of free parameters optimized; and  $n$  is the number of choices the subject made. The BIC was calculated based on the hierarchical model fitting. Note that since participants might employ either model-free or model-based strategy in individual trials, we did not choose RL model based on aggregated BIC of the participants.

### ***Magnetic Resonance Imaging (MRI) Images Acquisition***

Neuroimaging data were collected using a 3T whole-body Siemens MAGNETOM Skyra scanner with 64 channel head coil located at the Klinikum der Universität München. T2\* echo-planar images (with a multiband factor of 4) were collected with repetition time (TR) = 1760 ms, a matrix size of 106 x 106, an echo time (TE) = 38.6 ms, 45 deg flip angle, and field of view (FOV) = 212 x 212 mm. 60 axial slices were included in every volume. The voxel resolution is 2.0 x 2.0 x 2.0 mm<sup>3</sup>. The phase encoding direction was from anterior to posterior. Five dummy scans at the beginning of each run were discarded to allow for stabilization of the MR signal. The structural brain images were acquired applying an MP-RAGE T1-weighted sequence (TR/TE = 2.4 / 2.17 ms) with 320 x 256 matrix size, resolution 0.75 x 0.75 x 0.75 mm<sup>3</sup>, 12 deg flip angle, and a FOV of 240 x 240 mm.

### ***Preprocessing of the fMRI Data***

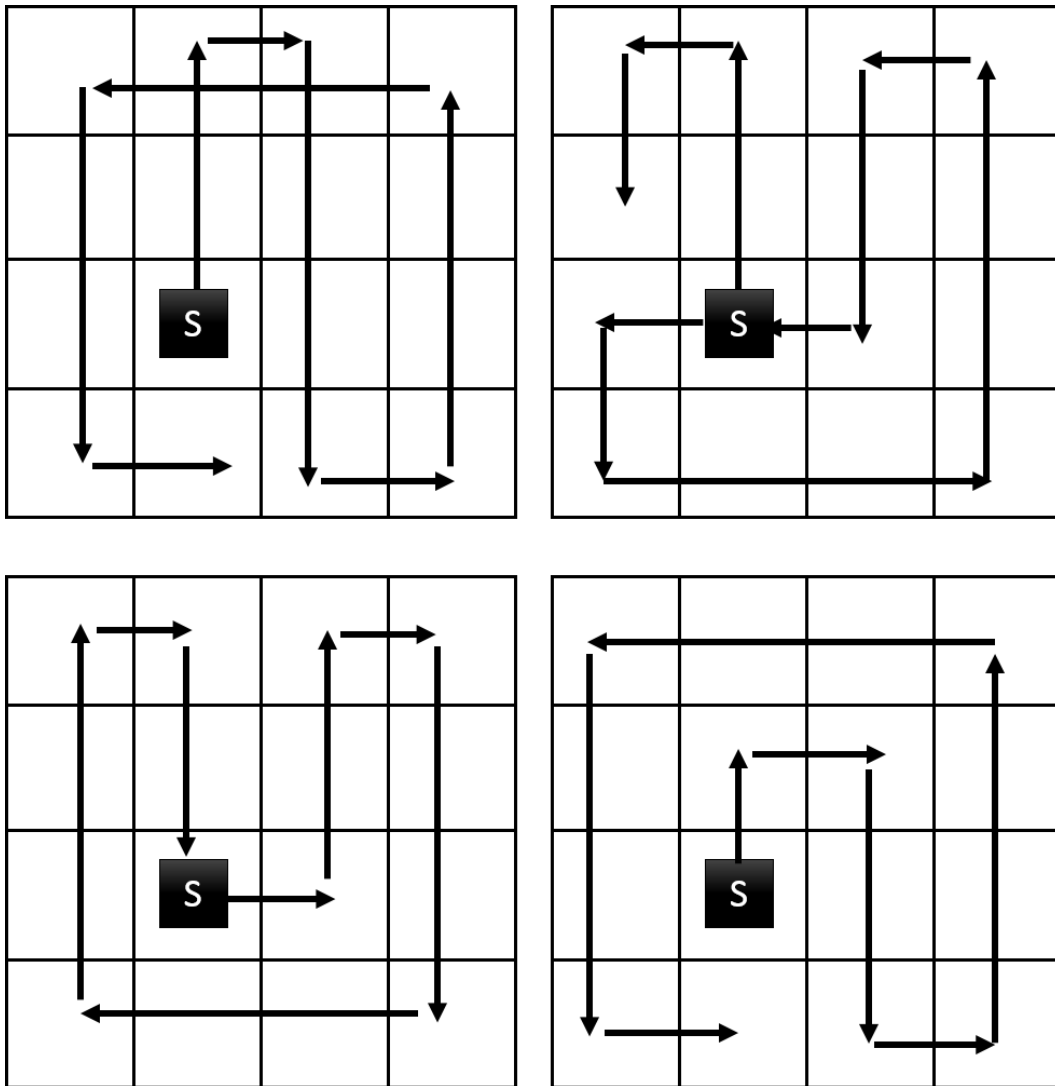
Neuroimaging data were processed and analysed using SPM12 toolbox ([www.fil.ion.ucl.ac.uk/spm/software/spm12/](http://www.fil.ion.ucl.ac.uk/spm/software/spm12/)). An event-related statistical analysis was applied to the images on two levels using the general linear model approach as implemented in SPM12.

Functional images were realigned for head motion and coregistered between runs and to the structural images. The images were then spatially normalized to Montreal Neurological Institute (MNI) space using the normalization parameters generated during the



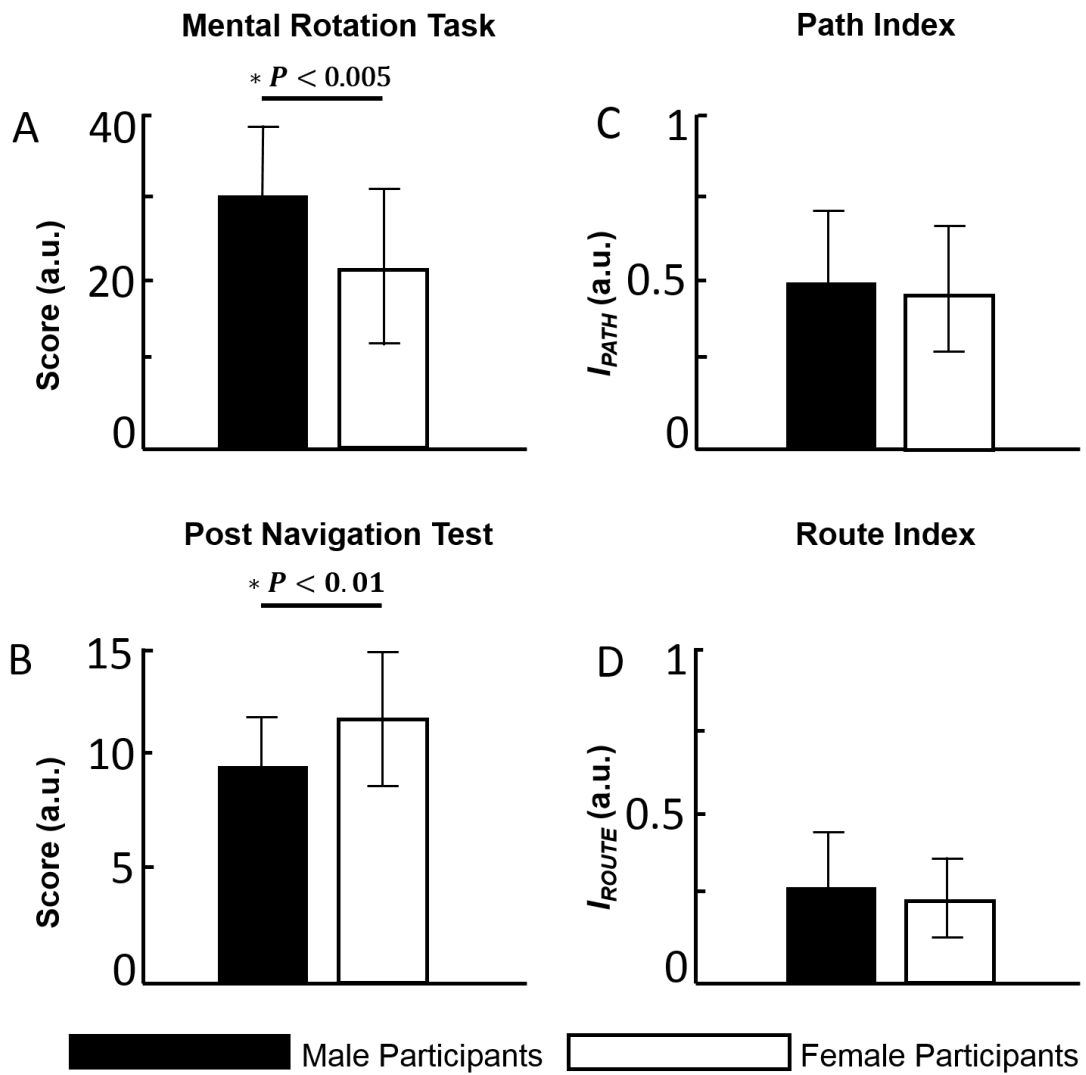
segmentation for each subject's anatomical T1 scan and resampled to 2 mm isotropic voxels. Subsequently, all images were smoothed with an 8 mm full width at half maximum Gaussian kernel.

## 3.8.3 Supplementary Figures



**Supplementary Figure 3.1 Path during the training phase**

During the training phase, participants were followed the instructions appearing on the screen to go through 4 different paths. Instructions indicated which door to take in each decision point, and allowed the participants to visit all the rooms in the grid-world. The paths followed during this training phase were the same for all participants. These four paths did not correspond to the paths the participants should take to retrieve the rewards. Rewards were hidden from participants' plain sight.



Supplementary Figure 3.2. Influence of gender on spatial ability

A. Significant difference between the two group results, with females scoring higher than males.

B. Significant difference between the two group results, with females scoring higher than males.

No significant differences were found among the two groups in either the C. optimal performance ratio ( $P = 0.133$ ) or D. the repeated runs ratio ( $P = 0.529$ ).

## 3.8.4 Supplementary Tables

Supplementary Table 3.1. Navigation Indices, MMSE Score, Fitted Parameter Weight, and MRT Score of the Older Participants

Subject	$I_{PATH}$	$I_{STEPS}$	MMSE	Weight ( $\omega$ )				Aggregate BIC			MRT
				Encoding	Retrieval	Search	Average	MF	MB	Hybrid	
1	0.17	0.35	28	0.03	0.05	0.49	0.19	252.2	290.3	259.4	12
2	0.04	0.43	27	0.02	0.03	0.03	0.03	196.1	241.5	203.2	9
3	0.11	0.49	28	0.02	0.07	1.00	0.36	214.5	262.0	224.3	13
4	0.35	0.42	27	0.15	0.53	0.70	0.46	263.4	308.9	268.4	7
5	0.11	0.39	28	0.03	0	0.08	0.04	195.5	220.9	202.2	8
6	0.02	0.44	27	0.02	0	0.03	0.02	162.0	218.5	168.8	7
7	0.13	0.38	28	0.39	0	0.73	0.37	236.7	259.3	242.0	14
8	0.13	0.41	29	0.46	0	0.03	0.17	330.2	332.8	336.9	8
9	0.22	0.36	26	0.03	0.02	0.61	0.22	425.3	431.7	433.5	8
10	0.16	0.37	28	0.28	0.02	1.00	0.43	299.4	307.4	305.3	12
11	0.07	0.38	26	0.19	0.32	0.69	0.40	273.7	300.7	280.6	11
12	0.02	0.62	27	0	0.16	0.61	0.26	172.8	283.2	177.8	9
13	0.07	0.46	27	0	0	0.61	0.20	240.9	282.9	247.8	9
14	0.09	0.56	27	0	0	0.44	0.15	138.3	228.5	144.8	8
15	0.07	0.54	26	0.02	0.14	0.46	0.21	115.4	250.4	121.4	10
16	0.07	0.48	25	0.02	0.03	0.68	0.24	103.4	218.3	106.3	12
17	0.22	0.48	28	0.03	0.05	0.03	0.04	198.4	310.2	206.1	15
18	0.23	0.42	25	0.22	0.20	0.67	0.36	284.9	309.1	290.7	13
19	0.18	0.46	26	0	0.34	0.03	0.13	338.3	358.4	346.4	14
20	0.18	0.57	25	0.50	0.10	0.62	0.41	185.3	251.0	187.1	12
21	0.17	0.36	26	0.80	0.02	0.82	0.55	405.6	390.3	412.7	10
22	0.17	0.38	25	0.05	0.03	0.74	0.27	357.8	354.0	365.3	11

Supplementary Table 3.2. Navigation Indices, MMSE Score, Fitted Parameter Weight, and MRT Score of the Female Younger Participants

Subject	$I_{PATH}$	$I_{STEPS}$	MMSE	Weight ( $\omega$ )				Aggregate BIC			MRT
				Encoding	Retrieval	Search	Average	MF	MB	Hybrid	
1	0.50	0.31	27	0.32	0.30	0.63	0.42	158.9	153.8	157.6	18
2	0.47	0.26	28	0.07	0.77	0.87	0.57	187.8	193.5	184.8	15
3	0.81	0.38	27	0.10	0.76	1.00	0.62	116.3	113.9	117.3	44
4	0.19	0.06	28	0.05	0.07	0.28	0.13	429.6	425.0	437.5	17
5	0.56	0.22	26	0	0.48	0.06	0.18	218.2	224.6	221.8	11
6	0.22	0.08	27	0.05	0.07	1.00	0.37	469.6	459.1	477.5	26
7	0.22	0.36	27	0	0.07	0.03	0.04	121.2	204.9	128.2	43
8	0.36	0.26	28	0	0.31	0.70	0.34	219.2	242.6	223.0	25
9	0.88	0.39	28	0.69	0.83	0.87	0.80	89.0	75.3	77.2	31
10	0.42	0.17	26	0.28	0.65	0.41	0.45	217.3	220.3	220.7	15
11	0.22	0.22	27	0.02	0.07	0.03	0.04	289.2	300.6	296.3	13
12	0.42	0.19	28	0.03	0.07	0.66	0.26	353.7	343.7	360.5	14
13	0.22	0.06	28	0.01	0.51	0.66	0.39	400.9	405.9	406.9	27
14	0.22	0.32	27	0.06	0.80	0.03	0.30	148.0	192.5	139.0	17
15	0.33	0.29	26	0.01	0.46	0.80	0.42	209.5	217.4	213.4	21
16	0.33	0.17	28	0.03	0.07	0.03	0.05	353.5	353.8	360.6	17
17	0.14	0.22	28	0.06	0.38	0.35	0.26	331.4	405.7	337.0	12
18	0.58	0.29	27	0.84	0.87	1.00	0.90	154.9	142.2	147.4	26
19	0.44	0.36	28	0.16	0.24	0.52	0.31	164.1	171.1	167.3	23
20	0.31	0.24	28	0.02	0.46	0.44	0.31	234.0	256.9	238.8	21

Supplementary Table 3.3. Navigation Indices, MMSE Score, Fitted Parameter Weight, and MRT Score of the Male Younger Participants

Subject	$I_{PATH}$	$I_{STEPS}$	MMSE	Weight ( $w$ )				Aggregate BIC			MRT
				Encoding	Retrieval	Search	Average	MF	MB	Hybrid	
1	0.08	0.33	27	0.03	0.05	0.49	0.19	234.7	282.8	242.0	29
2	0.38	0.24	27	0.02	0.03	0.03	0.03	168.9	157.9	169.2	24
3	0.56	0.14	28	0.02	0.07	1.00	0.36	157.6	157.5	149.1	23
4	0.56	0.31	28	0.15	0.53	0.70	0.46	135.0	143.8	140.0	30
5	0.33	0.19	27	0.03	0.00	0.08	0.04	246.2	258.3	254.6	14
6	0.50	0.50	26	0.02	0.00	0.03	0.02	131.2	150.0	133.8	32
7	0.22	0.19	26	0.39	0.00	0.73	0.37	306.4	311.5	313.7	25
8	0.17	0.06	27	0.46	0.00	0.03	0.17	393.8	388.9	402.3	28
9	0.44	0.22	28	0.03	0.02	0.61	0.22	230.8	225.0	233.4	28
10	0.25	0.11	27	0.28	0.02	1.00	0.43	400.7	394.3	408.8	19
11	0.76	0.40	27	0.19	0.32	0.69	0.40	90.6	81.7	81.5	33
12	0.40	0.17	28	0.00	0.16	0.61	0.26	345.4	335.1	350.6	23
13	0.78	0.38	26	0.00	0.00	0.61	0.20	96.5	86.0	90.2	25
14	0.68	0.22	28	0.00	0.00	0.44	0.15	123.0	102.1	107.2	36
15	0.85	0.38	28	0.02	0.14	0.46	0.21	89.5	78.9	81.8	40
16	0.58	0.31	27	0.02	0.03	0.68	0.24	122.6	146.1	126.1	14
17	0.28	0.14	27	0.03	0.05	0.03	0.04	401.2	404.0	410.8	28
18	0.31	0.17	28	0.22	0.20	0.67	0.36	256.7	268.1	260.7	40
19	0.57	0.32	27	0.00	0.34	0.03	0.13	120.0	106.2	109.2	44
20	0.76	0.35	28	0.50	0.10	0.62	0.41	98.4	89.4	89.0	39
21	0.81	0.51	28	0.80	0.02	0.82	0.55	86.8	100.0	91.4	46
22	0.58	0.19	28	0.05	0.03	0.74	0.27	149.6	131.6	144.1	35

Supplementary Table 3.4. Results of Model Fitting for the Older Participants

	Model-free		Model-based				Hybrid		Random			
<b>Encoding Phase</b>												
$\beta$	1.79	<b>2.60</b>	4.11	1.26	<b>3.51</b>	4.12	2.32	<b>3.04</b>	5.02			
$\gamma$				-	<b>0.1</b>	-	-	<b>0.1</b>	-			
$\alpha$	0.86	<b>1</b>	1				0.80	<b>1</b>	1			
$\lambda$	0.93	<b>0.98</b>	1				0.94	<b>1</b>	1			
$\omega$							0	<b>0.05</b>	0.21			
NLL	23.84	<b>67.14</b>	87.29	74.04	<b>97.04</b>	130.07	23.87	<b>66.88</b>	87.33	80.71	<b>97.21</b>	141.32
BIC	30.93	<b>74.51</b>	95.06	77.40	<b>99.68</b>	132.70	33.32	<b>76.71</b>	97.70	80.71	<b>97.21</b>	141.32
<b>Retrieval Phase</b>												
$\beta$	1.01	<b>1.63</b>	3.57	1.64	<b>1.83</b>	3.23	1.09	<b>1.88</b>	3.88			
$\gamma$				-	<b>0.1</b>	-	-	<b>0.1</b>	-			
$\alpha$	0.76	<b>1</b>	1				0.73	<b>1</b>	1			
$\lambda$	0.25	<b>0.95</b>	1				0.27	<b>0.92</b>	1			
$\omega$							0.02	<b>0.05</b>	0.07			
NLL	36.60	<b>62.80</b>	74.85	67.89	<b>75.91</b>	85.81	34.80	<b>62.22</b>	75.18	71.39	<b>85.44</b>	88.70
BIC	43.28	<b>69.31</b>	81.56	70.05	<b>78.12</b>	87.81	43.71	<b>70.90</b>	84.17	71.39	<b>85.44</b>	88.70
<b>Search Phase</b>												
$\beta$	1.35	<b>1.61</b>	2.37	1.43	<b>1.54</b>	1.65	1.35	<b>2.0</b>	3.39			
$\gamma$							-	<b>0.1</b>	-			
$\alpha$	0.1	<b>0.14</b>	0.80				0.02	<b>0.13</b>	0.82			
$\lambda$	0.07	<b>0.20</b>	0.71				0.05	<b>0.20</b>	0.62			
$\omega$							0.17	<b>0.61</b>	0.70			
NLL	66.73	<b>98.71</b>	109.72	79.28	<b>99.91</b>	115.32	65.68	<b>97.91</b>	74.96	82.91	<b>115.21</b>	132.06
BIC	73.69	<b>106.11</b>	117.27	81.68	<b>102.38</b>	117.87	74.96	<b>107.93</b>	120.31	82.91	<b>115.21</b>	132.06
$\rho^2$	0.18	<b>0.29</b>	0.45	0.11	<b>0.32</b>	0.39	0.16	0.28	0.31			

Supplementary Table 3.5. Results of the RL Model Fitting for the Male Younger Participants

	Model-free		Model-based				Hybrid		Random			
<b>Encoding Phase</b>												
$\beta$	2.14	<b>4.17</b>	20.0	0.76	<b>2.97</b>	8.77	2.34	<b>7.29</b>	17.30			
$\gamma$				-	<b>0.1</b>	-	-	<b>0.1</b>	-			
$\alpha$	0.08	<b>0.58</b>	1.0				0.10	<b>0.54</b>	1.0			
$\lambda$	0.41	<b>0.75</b>	0.89				0.56	<b>0.83</b>	0.96			
$\omega$							0.05	<b>0.06</b>	0.50			
NLL	55.24	<b>97.21</b>	156.42	62.62	<b>108.61</b>	162.67	49.75	<b>96.40</b>	156.37	106.29	<b>131.74</b>	226.12
BIC	62.11	<b>104.73</b>	164.55	64.93	<b>111.13</b>	165.38	58.70	<b>106.35</b>	167.20	106.29	<b>131.74</b>	226.12
<b>Retrieval Phase</b>												
$\beta$	2.43	<b>3.39</b>	6.19	3.01	<b>7.56</b>	15.99	3.18	<b>5.65</b>	16.32			
$\gamma$				-	<b>0.1</b>	-	-	<b>0.1</b>	-			
$\alpha$	0.47	<b>1.0</b>	1.0				0.43	<b>1.0</b>	1.0			
$\lambda$	0.55	<b>0.74</b>	1.0				0.58	<b>0.84</b>	1.0			
$\omega$							0.07	<b>0.47</b>	0.79			
NLL	14.40	<b>24.03</b>	42.93	18.42	<b>33.67</b>	59.25	12.14	<b>23.95</b>	40.96	52.49	<b>57.74</b>	67.52
BIC	18.22	<b>28.97</b>	49.11	19.69	<b>35.42</b>	61.31	18.09	<b>30.53</b>	48.94	52.49	<b>57.74</b>	67.52
<b>Search Phase</b>												
$\beta$	1.02	<b>1.83</b>	2.69	2.12	<b>5.64</b>	9.29	1.87	<b>3.34</b>	7.26			
$\gamma$							-	<b>0.1</b>	-			
$\alpha$	0.42	<b>1.0</b>	1.0				0.37	<b>0.96</b>	1.0			
$\lambda$	0.39	<b>0.76</b>	0.92				0.27	<b>0.64</b>	0.98			
$\omega$							0.08	<b>0.51</b>	0.85			
NLL	29.88	<b>46.71</b>	75.37	32.47	<b>48.79</b>	78.49	27.28	<b>45.53</b>	74.79	96.65	<b>114.38</b>	136.26
BIC	35.98	<b>52.92</b>	82.53	34.42	<b>50.93</b>	80.88	35.36	<b>52.0</b>	84.34	96.65	<b>114.38</b>	136.26
$\rho^2$	0.27	<b>0.36</b>	0.45	0.21	<b>0.37</b>	0.42	0.22	0.42	0.48			



**Supplementary Table 3.6. Activation at decision points for model-free valuation across three phases of the wayfinding task**

Regions	Peak Coordinates			Hemi	$T_{PEAK}$	k
	x	y	z			
Medial PFC						
Superior medial gyrus	6	42	40	R	10.35	3559
Medial frontal gyrus	-6	28	-10	L	6.11	132
Anterior Cingulate Cortex	4	34	20	R	8.78	3559
Middle Temporal Gyrus	48	-40	4	R	8.64	3886
	54	-24	-8	R	7.3	3886
	-56	-56	24	L	10.64	6052
Inferior Frontal Gyrus	52	18	6	R	8.64	2520
Cuneus	-4	-90	22	L	10.06	4729
Retrosplenial Complex	4	-24	38	R	9.23	4729
Precuneus	-6	-58	28	L	8.37	4729

Regions showing correlated BOLD activity with model-free values as parametric regressors across three different phases of the wayfinding task. Statistical significance was determined at the group level using a random-effect analysis. Regions listed exhibited significant peak voxels at probability threshold of  $P < 0.05$  based on a FWE cluster level small volume correction ( $k = 25$ ). Peak voxel MNI coordinates x, y, z are given in millimeters. L, left; R, right.

### 3.8.5 Supplementary References

- Caissie, A. F., Vigneau, F. & Bors, D. A. What does the Mental Rotation Test Measure? An Analysis of Item Difficulty and Item Characteristics. *The Open Psychology Journal* 2, 8 (2009).
- Harris, M. A., Wiener, J. M. & Wolbers, T. Aging specifically impairs switching to an allocentric navigational strategy. *Front Aging Neurosci* 4, 29, doi:10.3389/fnagi.2012.00029 (2012).
- Latini-Corazzini, L. et al. Route and survey processing of topographical memory during navigation. *Psychol Res* 74, 545-559, doi:10.1007/s00426-010-0276-5 (2010).
- Shikauchi, Y. & Ishii, S. Decoding the view expectation during learned maze navigation from human fronto-parietal network. *Scientific Reports* 5, 17648, doi:10.1038/srep17648 (2015).
- Vandenberg, S. G. & Kuse, A. R. Mental rotations, a group test of three-dimensional spatial visualization. *Percept Mot Skills* 47, 599-604, doi:10.2466/pms.1978.47.2.599 (1978).
- Viard, A., Doeller, C. F., Hartley, T., Bird, C. M. & Burgess, N. Anterior hippocampus and goal-directed spatial decision making. *J Neurosci* 31, 4613-4621, doi:10.1523/jneurosci.4640-10.2011 (2011).
- Voermans, N. C. et al. Interaction between the human hippocampus and the caudate nucleus during route recognition. *Neuron* 43, 427-435, doi:10.1016/j.neuron.2004.07.009 (2004).
- Wiener, J. M., de Condappa, O., Harris, M. A. & Wolbers, T. Maladaptive bias for extrahippocampal navigation strategies in aging humans. *J Neurosci* 33, 6012-6017, doi:10.1523/jneurosci.0717-12.2013 (2013).
- Wunderlich, K., Smittenaar, P. & Dolan, R. J. Dopamine enhances model-based over model-free choice behavior. *Neuron* 75, 418-424, doi:10.1016/j.neuron.2012.03.042 (2012).





# 4

## General Discussion

This chapter discusses the relevance of the findings of the thesis, and places these findings in relation to the existing literatures. In this thesis, the principal hypotheses were tested through analysis of empirical data that addressed the following research questions:

- (1) the utility of using reinforcement learning (RL) models to account for strategy adoption during spatial navigation,
- (2) the role of different brain regions in processing navigational decision by showing a correlation between BOLD activity and key variables of RL models,
- (3) the effects of normal aging on strategy adoption and its neural correlates.

All in all, this chapter highlights the merit of combining RL, VR-based wayfinding task, and fMRI as a way of bridging the gap between two rather separate fields: value-based decision making and spatial navigation. More importantly, this thesis opens new avenues for computational models of spatial navigation and wayfinding by directing attention to decisions, differing from common directions and distances approaches.

## 4.1 Summary of Performed Studies and Main Findings

In the first study (Chapter 2 – Manuscript 1), 27-right handed females (between 20 to 29 years of age) performed a VR-based wayfinding task while undergoing functional Magnetic Resonance Imaging (fMRI). The core of the task was a 5-by-5 grid of connected rooms where three different target objects, i.e. rewards, were to be found. To find the rewards, subjects could rely on route-based navigation, map-based navigation, or combination of both. We then modeled subjects' traversed paths using reinforcement learning (RL) algorithms. These algorithms have successfully explained decision behavior and its neural correlates. We found that subjects used a mixture of route and map-based navigation across trials. Their paths could be well explained by a combination of model-free and model-based RL algorithms. Using key internal variables from the RL algorithms, the model-free and model-based values, we aimed to elucidate the computational mechanism accounting for neurobiological underpinnings behind route and map-based navigation. Consistent with previous findings from both spatial navigation and decision literatures, the BOLD signals in the ventromedial prefrontal cortex (vmPFC) pertained to model-free value signals. In contrast, the computed value signals of model-based choices the BOLD signals in the medial temporal lobe (MTL). Based on these findings, we argue that the brain might share computational mechanisms and neural substrates for navigation and for making value-based decisions.

The second experiment (Chapter 3 – Manuscript 2) was a cross sectional aging study investigating the age-related shift in strategy adoption during spatial navigation in healthy adults. In total, 42 younger adults and 25 older adults (data from 3 older adults were not included in the analysis) participated in the study. We employed similar approach to the first study. However, we used a smaller grid world for the wayfinding task (4-by-4 instead of 5-by-5), and only the older participants performed the task inside the MRI scanner. As underlined by a growing number of studies, there are shifts from map-based navigation to route-based navigation with increasing age (Rodgers et al., 2012; Wiener et al., 2012). Some studies even demonstrated reduced or absent hippocampal/parahippocampal activation in the older adults as compared to the younger adults (Atonova et al., 2009; Moffat et al., 2006). In line with these findings, we found that older participants travelled longer distances to reach rewards. They also had a higher propensity to repeat previously established routes to locate target objects. Replicating the findings from the first study, the

traversed paths in both groups could be well explained by the model-free and model-based RL algorithms. Importantly, neuroimaging results from the older participants show that BOLD signals in the vmPFC pertained to model-free value signals. These results provide evidence that model-free RL explain how the aging brain computationally prefer to rely more on route-based navigation.

## 4.2 A Framework for Studying the Computational Mechanism of Route and Map-based Navigation

The core idea of this thesis is a computational framework for studying route and map-based navigation. We developed this framework by combining a novel wayfinding paradigm, computational models based on RL, and fMRI. Our approach provided a more flexible quantitative assessment than many other existing paradigms (Section 1.3.2). The reason for this is that our approach allowed us to measure one strategy preferentially compared with the other. There is one common approach often taken in the human spatial navigation literature, which we did not implement in this thesis. This approach, as mentioned in the beginning of this thesis (see introductory section of Chapter 1), is to dissociate route and map-based navigation by selectively studying several factors that may be unique to either type of navigation strategy. These factors include how information is acquired (route knowledge vs. survey knowledge), the context of the experimental task (navigating, pointing, drawing a map), or the reliability of the spatial information that has been acquired and integrated into the cognitive map. Although these works put the best effort to specifically design the task to be ‘purely’ route-based or ‘purely’ egocentric or ‘purely’ allocentric, during active navigation, no wayfinding task can guarantee a strict dichotomy of route and map-based navigation. It can, however, tip the spectrum toward one strategy, one spatial knowledge, or one reference frame.

Motivated by previous works in navigation research, we designed our wayfinding task without putting a restriction on which strategy subjects needed to use in which trial or phase of the experiment. As a measure of strategy preference, we used navigation indices and parameter  $\omega$  from RL models. These two behavioral measures allow the measure of strategy preference to be a continuum along which individuals may vary. As presented throughout this thesis and summarized in Section 4.1, our approach yielded promising results. In the following two subsections, we will discuss how our approach contributes to the existing knowledge. Specifically, on how computational mechanisms behind model-free and model-based choices may explain how the human brain encodes, updates, and uses spatial information. These subsections are organized based on cognitive processes pertaining to route-based and map-based navigation.



### 4.2.1 Cognitive Processes Pertaining to Route-based Navigation

There are three cognitive processes that are crucial to and the highlights of the route-based navigation. These processes include place recognition, sequence and response learning, as well as forming association.

#### *Place recognition – Scene and view*

During wayfinding, we must know our location in order to take the appropriate action. Depending on the navigator's spatial knowledge for a given environment, place recognition can have different implications (Philbeck and Sargent, 2012). One can use a scene or view-based system to recognize a unique place, which may only require object-place knowledge. As explained in details in Subsection 1.1.2, object-place and route knowledge do not require additional spatial localization within the environment.

In this thesis, we argue that our participants used scene-based place recognition when they recognized their current position based on the unique furniture in the room. Drawing a parallel to cognitive processes in value-based decision making (Subsection 1.5.1), place recognition is comparable to identifying external states. This is because, using this scene and view-based place recognition, participants could indicate 'I am at the living room'. In day to day navigation, this process translates to the ability to indicate 'I am at the train station' or 'I am at the bank'. Once participants recognized their current location (i.e. external states), they could identify which doors were available to choose from (i.e. possible course of actions).

Moreover, the results of post-navigation task (Manuscript 2 – Subsection 3.3.1, Figure 3.3) showed that older participants relied more on this scene-based cognitive processes. They scored significantly better on the post-navigation task compared to the younger cohorts. This is an indication that they had better object-place knowledge, and may rely more on scene-based place recognition in solving the wayfinding task. In line with this result, fitted parameter  $\omega$  also indicated that, on average, older adults showed stronger propensity towards model-free choice.

### ***Sequence and response learning***

Sequence learning is an important cognitive process in route-based navigation, because the ordinal relationship between landmarks provides important clues about where to turn, the proximity to goal, and importantly monitoring for errors. In our experiment, this cognitive process guides participants to encode ‘down this path, first I will encounter the dining room, then the living room, and then the wine cellar’. If the next expected landmark in a sequence does not appear, it is possible that the participants made a wrong turn at the previous location. Using this discrepancy between expected landmark and the one they actually encountered after making a choice, they could then update their current location. We can see that monitoring for errors in sequence learning is somewhat comparable to updating the prediction error, which is a hallmark of model-free choices (for details on value-based decision making and model-free choices, see Subsection 1.5.1 and 1.5.2).

Response learning enables one to take the necessary action at a location. This cognitive process is often studied in contrast with place learning. This classification is related to the longstanding distinction in the animal learning literature based on a conception of multiple parallel systems for learning and memory (O’Keefe and Nadel, 1978), and the two systems may indeed involve distinct reinforcement learning mechanisms (Simon and Daw, 2011). In this thesis, we argue that response learning is the cognitive process that guided our participants to ‘choose left in the dining room’. In a day to day context, this translates to ‘turn left at the bank’. Response learning is, thus, less flexible and involves associating each place or view with a particulate response in order to follow a known path to find reward.

### ***Forming associations***

Forming associations, which is a key cognitive process of route-based navigation, occurs at many points in navigation tasks. However, it might be carried out with differing purposes and outcomes. Forming association between a place and an action is crucial to response learning (Taylor and Brunye, 2012). Moreover, this cognitive process is not only important for spatial navigation. In value-based decision making, it is well reported that model-free choice behavior is motivated by the onset of a stimulus that is associated with certain outcomes. These stimulus-responses patterns are acquired by virtue of the extent to which a particular response give rise to positive or negative reinforcement (Section 1.5, for detailed review see O’Doherty et al., 2017).

In the context of the experiments presented in this thesis, forming stimulus-response patterns enabled our participants to deliberate whether to ‘choose left in the dining room that would take them to the living room’ or ‘choose right that would take them to the wine cellar’. Importantly, our results showed that subjects who exhibited a tendency toward route-based navigation were better explained by the model-free RL. Thus, we argued that we provide quantitative evidence that route-based navigation and model-free choice share similar computational mechanism.

#### **4.2.2 Cognitive Processes Pertaining to Map-based Navigation**

There are two cognitive processes that are crucial to map-based navigation and may share computational mechanisms with model-based choice. These processes include place recognition within the larger environment and locating the goal.

##### ***Place recognition within the larger environment***

In contrast to place recognition in route-based navigation, place recognition in map-based navigation requires situating the location within the broader environmental knowledge (Chrastil, 2012). In this thesis, this cognitive process means the ability to indicate ‘I am in the living room, which is on the left side of the arena and is near the dining room’. In day to day navigation, this may translate to ‘I am at the train station, which is on the east side of the town and is near the central bank’.

As explained in Subsection 1.1.2, survey knowledge require some information about a location beyond simple recognition. In other words, map-based place recognition require the knowledge of how a certain place fits within the larger scope of the environment. This cognitive process explains navigational behaviors of our participants that could not be logically reduced to associative learning between specific stimuli and rewarded behavioral response, especially during the retrieval and search phases. We observed that, for example, participants whose choice behaviors were well explained by the model-based RL, would plan a direct path to reach rewards if the established route took them longer. We concluded that these participants have had access to spatial knowledge about the arena, akin to the spatial knowledge obtainable from a map.

***Locating the goal***

When traveling to a specific goal, one must have some knowledge of where the goal is located within the larger environment. Some might argue that this process is no different with place recognition. However, knowledge of goal location means spatially relating the goal's location to the present location. This cognitive process is most important to map-based navigation. In route-based navigation, one may intend to reach a certain goal. However, one does not need to know where the goal is located within the large environment. Instead, following a series of place-action (i.e. stimulus-response) patterns is sufficient to reach the desired location.

Going back to the value-based decision making, model-based choice is motivated and directed toward a specific outcome (details in Subsection 1.5.3). It may also be thought of as prospective in that it leverages internal model of the decision problem to flexibly revalue states and actions (i.e. relating the reward state to the present state). Similar to map-based navigation, model-based choice facilitates a highly flexible system. This means, allowing one to adapt to changes in the environment without having to resample environment contingencies directly (O'Doherty et al., 2017).

In addition to understanding where the goal is located relative to the current position, map-based navigation also involves cognitive processes that allow the transformation of spatial information into useable action. This spatial information can be derived either from survey knowledge or allocentric reference frame. This means, information containing the connections of location or metric information must be transformed to how a navigator must move, turn, or choose certain direction in order to locate the goal. Again, this is analogous to model-based choice in a way that it calculates values of all relevant states and actions to determine which type of action yield the best outcome (Dayan and Niv, 2008; Rangel et al., 2008). Our results, subjects who used shortest paths to reach reward showed propensity toward model-based RL, provided quantitative evidence that map-based navigation and model-based choice share similar computations.

### **4.3 Neural Correlates of Cognitive Processes Pertaining to Route and Map-based Navigation**

This thesis is now set for the focal point of the general discussion: a critical appraisal of the neural correlates of cognitive processes involved in spatial navigation, and of their relationship to the proposed reinforcement learning (RL) framework. This section is organized based on the cognitive processes regardless whether they are associated with route-based or map-based navigation. At the end of this section, it will become clear, even in the broader sense of navigation, how spatial navigation and value-based decision making share various neural substrates.

#### **4.3.1 Place Recognition**

For objects or landmarks to aid place recognition, they must first be processed by the perceptual systems. There are three regions of the human brain that have been implicated in this function on the basis of their strong fMRI response during viewing of landmarks. These regions include:

- (1) The parahippocampal place area (PPA), which is a functional region that includes the parahippocampal cortex and nearby fusiform gyrus along the collateral sulcus near the posterior parahippocampal/anterior lingual boundary.
- (2) The retrosplenial complex (RSC), located in the parietal-occipital sulcus (POS), posterior to and partially overlapping with BA 29/30 (details on RSC in Subsection 1.4.3).
- (3) The occipital place area (OPA), located in the dorsal occipital lobe near the transverse occipital sulcus. These regions were initially studied in terms of their strong activation to visual scenes.

However, recent work suggests that these regions may be involved in processing both scene-like and object-like landmarks (Epstein and Vass, 2014). Consequently, activation in these regions, as reported in this thesis, indicate their substantial roles in place recognition during our wayfinding task.

Starting with a PET study in 1997 by Maguire et al., the PPA appeared to be involved in view-based scene recognition. The landmark task induced PET activation of the PPA, but

not the hippocampus, as compared with baseline (Maguire et al., 1997). This area also showed increased fMRI activity for scenes and landmarks presented in spatial context relative to faces and objects presented in isolation (Epstein and Kanwisher, 1998). In line with these findings, single unit recordings in humans showed that cells in the parahippocampal gyrus are more likely to respond to a particular view than are the hippocampal cells (Ekstrom and Bookheimer, 2007).

Furthermore, as discussed in Section 4.2, place recognition does not stop at recognizing a unique place based on scene or view. An additional step in place recognition, which is crucial to map-based navigation, is situating the location within the broader environmental context. In other words, anchoring landmarks to the cognitive map. Recent findings in human neuroimaging have provided evidence that the PPA is responsible for general contextual learning and for forming associations between objects and their location (Bar et al., 2008; Aminoff et al., 2013). This contextual learning effect appear to be strongest at slow presentation rates. Thus, it is often speculated that mental imagery of spatial location may be behind some of the activity in the parahippocampal cortex.

The results of our first study (Subsection 2.3.3) showed BOLD signals pertaining to model-based values, but not model-free, along the parahippocampal gyrus extending to the hippocampus and medial temporal lobe (MTL). As our behavioral results suggest that subjects who exhibited tendency toward model-based choice were more likely to rely on map-based navigation. Thus, model-based computations in the parahippocampal and MTL area during our wayfinding task could be indicative of forming links between landmark identity and their location within the broader environmental context.

As briefly mentioned in Subsection 1.4.3, one of the earliest fMRI study show that RSC activation occurred during scene viewing and mental imagery (Park and Chun, 2009). In line with this finding, of the three place recognition regions, RSC appears to be particularly important for using environmental cues for anchoring the cognitive map. fMRI response to scenes in RSC is significantly increased when subjects attempt to recover the location or implied heading of the scene within the broader spatial environment. In other words, RSC activation is particularly stronger when participants use the scene to localize or orient themselves (Epstein et al., 2007). In our first study, we found BOLD signals in the RSC pertaining to both model-free and model-based values. In our second study, we replicated

the neural correlates of the model-free values. This indicates that RSC roles are not limited to either route-based or map-based navigation. Instead, computations within the RSC might be the key process involving the use of environmental cues to determine the orientation and displacement within the cognitive map. The role of RSC associated with both route and map-based navigation will be elaborated in Subsection 4.2.5

### **4.3.2 Sequence, Response, and Associative Learning**

All three cognitive processes involves forming associations between place or view and a particular response. The main objective is to follow a known route to a goal. Using radial arm maze (Subsection 1.3.2) in a desktop VR, Iaria et al. (2003) found that individuals who relied on spatial relations between objects and distal landmarks (i.e. place learning) had an increase BOLD signal in the hippocampus. On the contrary, those who relied on the response learning and forming association between cues and actions showed increased BOLD signal in the caudate nucleus. Anatomical analysis of the caudate nucleus also reported that sequence and response learners had a greater density in the caudate nucleus instead of the hippocampus (Bohbot et al., 2007). In line with these findings, in the study by Doller et al. (2008), participants who learned to navigate to a location based on actions at a local landmark, showed increased BOLD signals in the dorsal striatum.

Furthermore, using a dual-strategy paradigm (DSP, details in Subsection 1.3.2), Marchette et al. (2011) found that some participants preferred to use map-based navigation in a complex maze by taking shortcuts to a target. Others, however, chose to use the response learning. These participants simply followed a known route, which is the hallmark of route-based navigation. BOLD activity in the response learning group showed marked increase in the caudate nucleus, regardless of participants' accuracy in the task.

From the perspective of value-based decision making, evidence suggested that the dorsal striatum is critical for learning the stimulus-response associations. Tricomi et al. (2009) demonstrated a link between BOLD signal in the human posterior striatum and response learning. Using reinforce devaluation test, they showed that increasing activity in the human posterior striatum is a function of training and the emergence of habitual control (Tricomi et al., 2009). Moreover, more recent studies supported these results by showing that activity

in this area correlated with the value of over-trained actions (Wunderlich et al., 2012) or model-free value signals (Horga et al., 2015).

In our experiments, we found correlates of model-free value signals in the caudate nucleus. Along with our results, findings from spatial navigation and value-based decision making literature presented above show that stimulus-response associations in both domains activate the caudate nucleus. This has important implication. That is, this thesis provide direct evidence that spatial navigation and value-based decision share not only computational mechanisms accounting for different cognitive processes, but also their neural substrates.

### **4.3.3 Locating the Goal**

In many studies of map-based navigation, it is not trivial to determine which aspects of the task are specific to identifying the goal, and which relate the goal to the current location. Spiers and Maguire (2006) observed increased activity in certain brain areas, relative to a coasting baseline, during sections of experiment where participants planned path to the target location. These brain areas included bilateral hippocampus, left parahippocampal cortex, the retrosplenial cortex, and the lateral and medial prefrontal cortex. The authors suggested that these areas may be important for identifying the location of the goal as well as the general sequence of the planned path (Spiers and Maguire, 2006). It has also been reported that microstructural integrity of the hippocampus is correlated with faster times navigating to a goal (Iaria et al., 2008). In addition, Viard et al. (2011) also reported that hippocampal activity increased with proximity to the goal. In the posterior hippocampus, some cells even appear to selectively respond to the navigation's goal (Ekstrom et al., 2003; Viard et al., 2011). However, these results might pinpoint to three possible cognitive processes including (1) navigator's ability to relate the goal with current location, (2) transformation between allocentric and egocentric reference frame, and (3) the use of metric distance information.

Another aspect of identifying the goal's location is transforming spatial information from one reference frame to the other. In a route-learning task Wolbers et al. (2004) found increased BOLD signal in the retrosplenial cortex during encoding compared to baseline.



The level of neural activity in the retrosplenial cortex was found to be stable throughout the encoding session (Wolbers et al., 2004). In a follow up study, Wolbers and Büchel (2005) asked participants to determine the relative location of a target object relative to the current location. Interestingly, this time they found that retrosplenial cortex showed performance-related increases in activity with each encoding session. The more information participants acquired about the maze, the greater the retrosplenial activity. In other words, the retrosplenial cortex was sensitive to the amount of spatial knowledge (Wolbers and Büchel, 2005).

The findings regarding how the retrosplenial cortex relates to navigation are somewhat conflicting. This region appears to be active in both route-based (Ino et al., 2002) and map-based navigation (Galati et al., 2000; Epstein and Higgins, 2007; Epstein et al., 2007). As mentioned in Subsection 4.3.1, we also found correlates of BOLD signal pertaining to both model-free and model-based value signals. Based on previous findings, the modulated RSC activity by both model-based and model-free values may reflect the robust integration of different sources of spatial information. We speculate that this region may serve as a transition zone between inputs in egocentric and allocentric reference frames. This is also consistent with the idea that RSC allows the translation from ,you are here' information to ,your goal is to the left' information

## 4.4 Decision in Space: Where Do We Go from Here?

In previous sections, findings from the studies presented in this thesis are critically discussed in the context of initial hypotheses and existing literature. During the development of this research, it became clear that several new aspects may be considered. In addition, there are also several avenues of inquiry that could advance the fields considerably. Thus, the first two subsections outline some of the remaining gaps in the literature and suggest ways in which these gaps may be filled (Subsection 4.4.1 and 4.4.2).

### 4.4.1 Potential Research Constraints

The literatures reviewed in this thesis are largely consistent with the classical models of the neural basis of route-based and map-based navigation. In other words, there are some degrees of specialization in the brain regions that code spatial processing pertaining to either route-based or map-based navigation. The results presented in this thesis are also consistent with these existing literatures. However, the results of other studies reveal that in some cases, the route-map dichotomy is not so clear cut. For instance, there is now growing evidence for some degree of redundant, distributed coding across brain areas. Early findings from neuroimaging studies of spatial navigation was in fact reporting a range of different brain regions that activate in contrasts (Galati et al., 2000; Committeri et al., 2004). Additionally, there is also growing consensus that the hippocampus and MTL respond to a larger variety of variables than just location, including viewing landmarks, directions, and conjunctions of these variables with spatial location (Ekstrom et al., 2003; Ekstrom and Bookheimer, 2007). Altogether, instead of being the specialized focus of a single brain region, the pattern of literatures suggests a sharing of computations related to spatial processing across many brain regions. Our approach in this thesis, cannot yet account for this emerging concept.

Furthermore, in this thesis, key parameters of the RL model were fitted as a single value for each phase of the wayfinding task. For example, parameter  $\omega$  was assumed to be constant throughout the entire encoding phase. This was also true for the retrieval and search phases. This approach provide a snapshot of how individual subject integrate route and map-based navigation in solving the wayfinding task. It did not, however, elucidate how or whether individual subject switch strategy from one trial to the next. This would be especially interesting for the search phase, where subjects were forced to decide whether to plan a

novel path or trace a path to a familiar location and follow an established route from that familiar location.

The last potential research constraint concern with the cross sectional study of aging. Cross-sectional studies, comparing two groups of participants, of aging tend to produce large effect sizes. In addition, the results are relatively consistent and reproducible, as demonstrated by the results of the second study in this thesis (Subsection 3.3.3). Nevertheless, within the framework of cross-sectional study, it is impossible to discern the influence of life-long individual differences in the brain and cognitive variables from true longitudinal declines. Moreover, some studies also reported gender as a factor that may influence individual differences in spatial navigation. Cross-sectional studies in older populations may allow for a snapshot of main effect of gender, but cannot resolve the issue of whether men and women may manifest different rates of navigation decline.

These constraints raised ideas for future direction of research in spatial navigation that will be elaborated in the next subsection.

#### **4.4.2 Implication for the Future of Spatial Navigation Research**

There are a least three important implications that our results bring for the future of spatial navigation research. These include the importance of (1) looking at neural correlates as a network of brain regions instead of assigning a specific cognitive process to a specific brain region, (2) longitudinal studies in navigation research, and (3) clinical potential of spatial navigation because, in many cases, reports of impaired spatial behavior are a major trigger to the diagnosis.

##### ***Toward a network conceptualization of spatial navigation***

The reviewed literature along with the results presented in this thesis suggest that there is some specialization in the involvement of brain regions in different cognitive processes that are crucial for spatial navigation. Importantly, however, the same previous findings also provide a clear evidence for shared, distributed representations across brain regions. For example, the hippocampus, parahippocampal gyrus, and RSC all carry information about location, heading, and distance between locations. Altogether, the emerging pattern of

results suggests a sharing of computations related to spatial processing across many brain regions.

Moreover, this network conceptualization also allows a continuous form of route and map-based navigation. This continuous form may be viewed as the navigation counterpart of the hybrid algorithm of the RL model. This rather fluid interpretation of route and map-based dichotomy would otherwise be difficult to conceptualize in a scenario where one brain region pertains specifically to one cognitive process. Further work on the hybrid algorithm of the RL model, such as allowing the parameter  $\omega$  to vary from trial to trial, may better facilitate this network conceptualization. This way, we don't have to assume whether route-based navigation and model-free values or map-based navigation and model-based values are integrated at every trial or calculated separately. Both forms of navigation strategy as well as decision making strategies can simultaneously coexist and be called on to solve a task, depending on the current demands.

### ***Longitudinal studies***

When it comes to aging study, based on the constraints explained in the previous subsection, the first and most obvious need in the field is for longitudinal research. On the one hand, longitudinal research may elucidate individual differences in rates at which cognitive declines affect navigation. Additionally, longitudinal approach may resolve the debate on what constitutes of 'normal aging' participants. In most aging studies, including the one presented in this thesis, 'normal aging' is defined as older adults who are healthy, living in the community, and willing to participate in the experiment. More recently, however, there has been a push to put a distinction between 'usual' and 'successful' aging within the construct of normal aging. Those who are aging successfully suffer few if any cognitive problems. With cross-sectional studies, oftentimes these two subgroups of normal older adults are mixed in the same group resulting in higher group variance compared to the younger adults. In contrast, with longitudinal research, we can compare paired data from the same individuals as they age.

### ***Clinical potential of spatial navigation***

This thesis highlights the behavioral manifestations and putative neural substrates of strategy adoption in spatial navigation. When it comes to the effect of aging on spatial

navigation, this thesis focuses on normal aging – that is, cognitive and brain aging that occurs in the absence of diagnosed neuropathology. Having stated that, it is important to recognize that both neuroscientists and the lay public are intensely interested in pathological aging. Alzheimer’s disease (AD) and mild cognitive impairment (MCI) being the public concern when it comes to age-related cognitive deficits. A common, though currently underexplored aspect of AD is that of topographical disorientation. Even very early AD patients may become disoriented. The early emergence of topographical disorientation in AD is to be expected. This is because of the overlap of the neural substrates of spatial computation with AD pathology. This overlap includes the regions of MTL and RSC (Braak et al., 2006; Pengas et al., 2010). Motivated by this observation, we are proposing three potential aspects on how the combination of RL modeling, wayfinding task, and fMRI can be useful in the clinical domain.

To begin with, in this thesis, we showed that subjects’ tendency toward route or map-based navigation was well explained by the RL algorithms. Thus, key variables of RL models and how these variables correlate with BOLD signals may have the potential utility as biomarkers for certain age-related pathologies. This stems from observations and findings that reported substantial navigations impairment in AD and MCI in wayfinding tasks that require substantial translation from an allocentric reference frame to an egocentric reference frame (Serino and Riva, 2013). Moreover deficits in route-based navigation in early AD have been associated with hypometabolism in the RSC, thalamus, and parietal cortex (Pengas et al., 2012). Cumulatively, these findings suggests that changes in the navigation circuit may be a predominant and early consequences of AD. These changes manifest as behavioral deficiencies in spatial navigation. Since the combination of RL and fMRI analysis allow us to analyze how MRI signal changes correlate with quantitative measure of behavioral performance, this approach may prove to be particularly useful in developing a biomarkers for certain age-related pathologies.

A second aspect of potential clinical utility is using RL’s key variables and their correlates to neural data as predictors of future AD or MCI onset on individual basis. Going back to the arguments presented in the previous subsection, the importance of longitudinal studies in clinical context lies on their potential to elucidate the rate of conversion to AD among MCI patients. One such study (Laczo et al., 2010) demonstrated that MCI patients who progressed to AD showed higher deficits in both route and map-based navigation. Should

these deficits can be quantified using RL models and these trends also present on the neuronal level, this approach may prove useful for early diagnosis and intervention.

Lastly, aiding treatment and prevention. In this domain, the combination of wayfinding task with RL modeling and fMRI can be used both as an outcome variable and as an intervention. As an outcome variable, this approach may be useful to assess the efficacy of clinical or pharmacological intervention. In addition, behavioral interventions in the form of cognitive training may enhance cognitive performance in healthy elderly. These interventions may also delay the onset or ameliorate the symptoms of dementia.

In the foreseeable future, we expect that more researchers and clinicians will incorporate wayfinding tasks into clinical research and practice. However, the challenges remain to establish the gold standard of wayfinding task to be used in clinical settings. Such task needs to be quick to administer, requires minimal training, and meet quality criteria for diagnostic tests.

## 4.5 Conclusion

From finding one's car in a parking lot to getting home from work, from finding one's cubicle in a new office building to finding a subway stop in an unfamiliar location, successful navigation requires the ability to identify and adopt the most suitable wayfinding strategy according to the task's demands. Strategy use, however, varies among individuals and changes significantly with ages. Some people are adept at flexibly change or integrate different wayfinding strategies. Others, such as older adults, experience increasing difficulties with the use of map-based navigation. This thesis contributes to a body of research that brings together spatial cognition and space syntax approaches. It offers a novel methodology for examining individual as well as age-related differences in strategy use. Importantly, it proposes a computational methodology for identifying brain areas that are crucial for successful navigation.









# References

- Adams S, Kesner RP, Ragozzino ME (2001) Role of the medial and lateral caudate-putamen in mediating an auditory conditional response association. *Neurobiology of learning and memory* 76:106-116.
- Aguirre GK, D'Esposito M (1997) Environmental knowledge is subserved by separable dorsal/ventral neural areas. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 17:2512-2518.
- Ainge JA, Tamosiunaite M, Woergoetter F, Dudchenko PA (2007) Hippocampal CA1 place cells encode intended destination on a maze with multiple choice points. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27:9769-9779.
- Alexander AS, Nitz DA (2015) Retrosplenial cortex maps the conjunction of internal and external spaces. *Nat Neurosci* 18:1143-1151.
- Alexander AS, Nitz DA (2017) Spatially Periodic Activation Patterns of Retrosplenial Cortex Encode Route Sub-spaces and Distance Traveled. *Current biology : CB* 27:1551-1560.e1554.
- Alexander GE, Crutcher MD, DeLong MR (1990) Basal ganglia-thalamocortical circuits: parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. *Progress in brain research* 85:119-146.
- Alfano PL, Michel GF (1990) Restricting the field of view: perceptual and performance effects. *Perceptual and motor skills* 70:35-45.
- Allen GL (2006) *Human Spatial Memory: Remembering Where*. New Jersey: Lawrence Erlbaum Associates.

- Amiez C, Joseph JP, Procyk E (2006) Reward encoding in the monkey anterior cingulate cortex. *Cerebral cortex* (New York, NY : 1991) 16:1040-1055.
- Aminoff EM, Kveraga K, Bar M (2013) The role of the parahippocampal cortex in cognition. *Trends in cognitive sciences* 17:379-390.
- Antonova E, Parslow D, Brammer M, Dawson GR, Jackson SH, Morris RG (2009) Age-related neural activity during allocentric spatial memory. *Memory* (Hove, England) 17:125-143.
- Appleyard D (1970) Styles and methods of structuring a city. *Environment and Behavior* 2:17.
- Astur RS, Taylor LB, Mamelak AN, Philpott L, Sutherland RJ (2002) Humans with hippocampus damage display severe spatial memory impairments in a virtual Morris water task. *Behavioural brain research* 132:77-84.
- Bakker NH, Passenier PO, Werkhoven PJ, Stassen HG, Wieringa PA (2001) Navigation in Virtual Environments. *IFAC Proceedings Volumes* 34:89-94.
- Bar M, Aminoff E, Schacter DL (2008) Scenes unseen: the parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28:8539-8544.
- Barracough DJ, Conroy ML, Lee D (2004) Prefrontal cortex and decision making in a mixed-strategy game. *Nat Neurosci* 7:404-410.
- Basten U, Biele G, Heekeren HR, Fiebach CJ (2010) How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Sciences of the United States of America* 107:21767-21772.
- Behrens TE, Woolrich MW, Walton ME, Rushworth MF (2007) Learning the value of information in an uncertain world. *Nat Neurosci* 10:1214-1221.
- Bohbot VD, Corkin S (2007) Posterior parahippocampal place learning in H.M. *Hippocampus* 17:863-872.
- Bohbot VD, Iaria G, Petrides M (2004) Hippocampal function and spatial memory: evidence from functional neuroimaging in healthy participants and performance of patients with medial temporal lobe resections. *Neuropsychology* 18:418-425.

- Bohbot VD, Lerch J, Thorndyraft B, Iaria G, Zijdenbos AP (2007) Gray matter differences correlate with spontaneous strategies in a human virtual navigation task. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27:10078-10083.
- Bohbot VD, Kalina M, Stepankova K, Spackova N, Petrides M, Nadel L (1998) Spatial memory deficits in patients with lesions to the right hippocampus and to the right parahippocampal cortex. *Neuropsychologia* 36:1217-1238.
- Boorman ED, Behrens TE, Woolrich MW, Rushworth MF (2009) How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron* 62:733-743.
- Bottini G, Cappa S, Geminiani G, Sterzi R (1990) Topographic disorientation--a case report. *Neuropsychologia* 28:309-312.
- Braak H, Alafuzoff I, Arzberger T, Kretschmar H, Del Tredici K (2006) Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta neuropathologica* 112:389-404.
- Burgess N, Maguire EA, O'Keefe J (2002) The human hippocampus and spatial and episodic memory. *Neuron* 35:625-641.
- Burles F, Slone E, Iaria G (2017) Dorso-medial and ventro-lateral functional specialization of the human retrosplenial complex in spatial updating and orienting. *Brain structure & function* 222:1481-1493.
- Carelli L, Rusconi ML, Scarabelli C, Stampatori C, Mattioli F, Riva G (2011) The transfer from survey (map-like) to route representations into Virtual Reality Mazes: effect of age and cerebral lesion. *Journal of neuroengineering and rehabilitation* 8:6.
- Carlen M (2017) What constitutes the prefrontal cortex? *Science (New York, NY)* 358:478-482.
- Chen LL, Lin LH, Barnes CA, McNaughton BL (1994a) Head-direction cells in the rat posterior cortex. II. Contributions of visual and ideothetic information to the directional firing. *Experimental brain research* 101:24-34.
- Chen LL, Lin LH, Green EJ, Barnes CA, McNaughton BL (1994b) Head-direction cells in the rat posterior cortex. I. Anatomical distribution and behavioral modulation. *Experimental brain research* 101:8-23.

- Chersi F, Burgess N (2015) The Cognitive Architecture of Spatial Navigation: Hippocampal and Striatal Contributions. *Neuron* 88:64-77.
- Chrastil ER (2013) Neural evidence supports a novel framework for spatial navigation. *Psychonomic bulletin & review* 20:208-227.
- Chowdhury R, Guitart-Masip M, Lambert C, Dayan P, Huys Q, Duzel E, Dolan RJ (2013) Dopamine restores reward prediction errors in old age. *Nat Neurosci* 16:648-653.
- Colombo PJ, Davis HP, Volpe BT (1989) Allocentric spatial and tactile memory impairments in rats with dorsal caudate lesions are affected by preoperative behavioral training. *Behavioral neuroscience* 103:1242-1250.
- Committeri G, Galati G, Paradis AL, Pizzamiglio L, Berthoz A, LeBihan D (2004) Reference frames for spatial cognition: different brain areas are involved in viewer-, object-, and landmark-centered judgments about object location. *Journal of cognitive neuroscience* 16:1517-1535.
- Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. *Nature* 441:876-879.
- Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69:1204-1215.
- Dayan P, Niv Y (2008) Reinforcement learning: the good, the bad and the ugly. *Current opinion in neurobiology* 18:185-196.
- De Leonibus E, Oliverio A, Mele A (2005) A study on the role of the dorsal striatum and the nucleus accumbens in allocentric and egocentric spatial memory consolidation. *Learning & memory (Cold Spring Harbor, NY)* 12:491-503.
- De Leonibus E, Mele A, Oliverio A, Pert A (2001) Locomotor activity induced by the non-competitive N-methyl-D-aspartate antagonist, MK-801: role of nucleus accumbens efferent pathways. *Neuroscience* 104:105-116.
- Delgado MR, Li J, Schiller D, Phelps EA (2008) The role of the striatum in aversive learning and aversive prediction errors. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 363:3787-3800.

- Diemer J, Alpers GW, Peperkorn HM, Shiban Y, Muhlberger A (2015) The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Frontiers in psychology* 6:26.
- Doeller CF, King JA, Burgess N (2008) Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *Proceedings of the National Academy of Sciences of the United States of America* 105:5915-5920.
- Dombeck DA, Reiser MB (2012) Real neuroscience in virtual worlds. *Current opinion in neurobiology* 22:3-10.
- Driscoll I, Hamilton DA, Yeo RA, Brooks WM, Sutherland RJ (2005) Virtual navigation in humans: the impact of age, sex, and hormones on place learning. *Hormones and behavior* 47:326-335.
- Duarte IC, Ferreira C, Marques J, Castelo-Branco M (2014) Anterior/posterior competitive deactivation/activation dichotomy in the human hippocampus as revealed by a 3D navigation task. *PloS one* 9:e86213.
- Ekstrom AD, Bookheimer SY (2007) Spatial and temporal episodic memory retrieval recruit dissociable functional networks in the human brain. *Learning & Memory* 14:645-654.
- Ekstrom AD, Arnold A, Iaria G (2014) A critical review of the allocentric spatial representation and its neural underpinnings: toward a network-based perspective. *Frontiers in human neuroscience* 8.
- Ekstrom AD, Huffman DJ, Starrett M (2017) Interacting networks of brain regions underlie human spatial navigation: a review and novel synthesis of the literature. *Journal of neurophysiology* 118:3328-3344.
- Ekstrom AD, Kahana MJ, Caplan JB, Fields TA, Isham EA, Newman EL, Fried I (2003) Cellular networks underlying human spatial navigation. *Nature* 425:184-188.
- Eppinger B, Hammerer D, Li SC (2011) Neuromodulation of reward-based learning and decision making in human aging. *Annals of the New York Academy of Sciences* 1235:1-17.
- Eppinger B, Schuck NW, Nystrom LE, Cohen JD (2013) Reduced striatal responses to reward prediction errors in older compared with younger adults. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33:9905-9912.

- Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. *Nature* 392:598-601.
- Epstein RA, Higgins JS (2007) Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cerebral cortex* (New York, NY : 1991) 17:1680-1693.
- Epstein RA, Vass LK (2014) Neural systems for landmark-based wayfinding in humans. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 369:20120533.
- Epstein RA, Parker WE, Feiler AM (2007) Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27:6141-6149.
- Ferbinteanu J, Shapiro ML (2003) Prospective and retrospective memory coding in the hippocampus. *Neuron* 40:1227-1239.
- Filimon F (2015) Are All Spatial Reference Frames Egocentric? Reinterpreting Evidence for Allocentric, Object-Centered, or World-Centered Reference Frames. *Frontiers in human neuroscience* 9:648.
- FitzGerald TH, Seymour B, Dolan RJ (2009) The role of human orbitofrontal cortex in value comparison for incommensurable objects. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29:8388-8395.
- Foo P, Warren WH, Duchon A, Tarr MJ (2005) Do humans integrate routes into a cognitive map? Map- versus landmark-based navigation of novel shortcuts. *Journal of experimental psychology Learning, memory, and cognition* 31:195-215.
- Foo P, Duchon A, Warren WH, Jr., Tarr MJ (2007) Humans do not switch between path knowledge and landmarks when learning a new environment. *Psychol Res* 71:240-251.
- Galati G, Lobel E, Vallar G, Berthoz A, Pizzamiglio L, Le Bihan D (2000) The neural basis of egocentric and allocentric coding of space in humans: a functional magnetic resonance study. *Experimental brain research* 133:156-164.
- Gazova I, Laco J, Rubinova E, Mokrisova I, Hyncicova E, Andel R, Vyhnalek M, Sheardova K, Coulson EJ, Hort J (2013) Spatial navigation in young versus older adults. *Frontiers in aging neuroscience* 5:94.



- Gillner S, Mallot HA (1998) Navigation and acquisition of spatial knowledge in a virtual maze. *Journal of cognitive neuroscience* 10:445-463.
- Glascher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66:585-595.
- Glascher J, Adolphs R, Damasio H, Bechara A, Rudrauf D, Calamia M, Paul LK, Tranel D (2012) Lesion mapping of cognitive control and value-based decision making in the prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America* 109:14681-14686.
- Glimcher PW, Rustichini A (2004) Neuroeconomics: the consilience of brain and decision. *Science (New York, NY)* 306:447-452.
- Goeke C, Kornpetpanee S, Koster M, Fernandez-Revelles AB, Gramann K, Konig P (2015) Cultural background shapes spatial reference frame proclivity. *Scientific reports* 5:11426.
- Golledge RG (1999) *Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes*. Baltimore: The John Hopkins University Press.
- Gomez A, Cerles M, Rousset S, Remy C, Baciú M (2014) Differential hippocampal and retrosplenial involvement in egocentric-updating, rotation, and allocentric processing during online spatial encoding: an fMRI study. *Frontiers in human neuroscience* 8:150.
- Grabenhorst F, Rolls ET, Margot C, da Silva MA, Velazco MI (2007) How pleasant and unpleasant stimuli combine in different brain regions: odor mixtures. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27:13532-13540.
- Greene KK, Donders J, Thoits T (2006) Topographical heading disorientation: a case study. *Applied neuropsychology* 13:269-274.
- Gron G, Wunderlich AP, Spitzer M, Tomczak R, Riepe MW (2000) Brain activation during human navigation: gender-different neural networks as substrate of performance. *Nat Neurosci* 3:404-408.

- Grubert J, Langlotz T, Zollmann S, Regenbrecht H (2017) Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality. *IEEE transactions on visualization and computer graphics* 23:1706-1724.
- Hammerer D, Eppinger B (2012) Dopaminergic and prefrontal contributions to reward-based learning and outcome monitoring during child development and aging. *Developmental psychology* 48:862-874.
- Hamilton DA, Akers KG, Johnson TE, Rice JP, Candelaria FT, Redhead ES (2009) Evidence for a shift from place navigation to directional responding in one variant of the Morris water task. *Journal of experimental psychology Animal behavior processes* 35:271-278.
- Hampton AN, Bossaerts P, O'Doherty JP (2006) The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 26:8360-8367.
- Hare TA, O'Doherty J, Camerer CF, Schultz W, Rangel A (2008) Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28:5623-5630.
- Harris MA, Wolbers T (2012) Ageing effects on path integration and landmark navigation. *Hippocampus* 22:1770-1780.
- Harris MA, Wiener JM, Wolbers T (2012) Aging specifically impairs switching to an allocentric navigational strategy. *Frontiers in aging neuroscience* 4:29.
- Hartley T, Maguire EA, Spiers HJ, Burgess N (2003) The well-worn route and the path less travelled. Distinct neural bases of route-following and wayfinding in humans. *Neuron* 37:877-888.
- Hayden BY, Platt ML (2010) Neurons in anterior cingulate cortex multiplex information about reward and action. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30:3339-3346.
- Heft H (2012) Wayfinding, Navigation, and Environmental Cognition from A Naturalist's Stance. In: *Handbook of Spatial Cognition* (Nadel L, Waller D, eds). Washington: American Psychological Association.

- Hegarty M, Waller D (2005) Individual differences in spatial abilities. In: Handbook of visuospatial thinking (Shah P, Miyake A, eds), pp 121-169. New York: Cambridge University Press.
- Herting MM, Nagel BJ (2012) Aerobic fitness relates to learning on a virtual Morris Water Task and hippocampal volume in adolescents. *Behavioural brain research* 233:517-525.
- Holden MP, Newcombe NS (2012) The Development of Location Coding: An Adaptive Combination Account. In: Handbook of Spatial Cognition (Nadel L, Waller D, eds). Washington: American Psychological Association.
- Horga G, Maia TV, Marsh R, Hao X, Xu D, Duan Y, Tau GZ, Graniello B, Wang Z, Kangarlou A, Martinez D, Packard MG, Peterson BS (2015) Changes in corticostriatal connectivity during reinforcement learning in humans. *Human brain mapping* 36:793-803.
- Huettel SA, McKeown MJ, Song AW, Hart S, Spencer DD, Allison T, McCarthy G (2004) Linking hemodynamic and electrophysiological measures of brain activity: evidence from functional MRI and intracranial field potentials. *Cerebral cortex* (New York, NY : 1991) 14:165-173.
- Hunt LT, Kolling N, Soltani A, Woolrich MW, Rushworth MF, Behrens TE (2012) Mechanisms underlying cortical activity during value-guided choice. *Nat Neurosci* 15:470-476, s471-473.
- Iaria G, Petrides M, Dagher A, Pike B, Bohbot VD (2003) Cognitive strategies dependent on the hippocampus and caudate nucleus in human navigation: variability and change with practice. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 23:5945-5952.
- Iaria G, Chen JK, Guariglia C, Püto A, Petrides M (2007) Retrosplenial and hippocampal brain regions in human navigation: complementary functional contributions to the formation and use of cognitive maps. *The European journal of neuroscience* 25:890-899.
- Iaria G, Lanyon LJ, Fox CJ, Giaschi D, Barton JJ (2008) Navigational skills correlate with hippocampal fractional anisotropy in humans. *Hippocampus* 18:335-339.

- Iglói K, Doeller CF, Berthoz A, Rondi-Reig L, Burgess N (2010) Lateralized human hippocampal activity predicts navigation based on sequence or place memory. *Proceedings of the National Academy of Sciences* 107:14466-14471.
- Ino T, Inoue Y, Kage M, Hirose S, Kimura T, Fukuyama H (2002) Mental navigation in humans is processed in the anterior bank of the parieto-occipital sulcus. *Neurosci Lett* 322:182-186.
- Ino T, Doi T, Hirose S, Kimura T, Ito J, Fukuyama H (2007) Directional disorientation following left retrosplenial hemorrhage: a case report with fMRI studies. *Cortex; a journal devoted to the study of the nervous system and behavior* 43:248-254.
- Jacob PY, Casali G, Spieser L, Page H, Overington D, Jeffery K (2017) An independent, landmark-dominated head-direction signal in dysgranular retrosplenial cortex. *Nat Neurosci* 20:173-175.
- Jacobs J, Kahana MJ, Ekstrom AD, Mollison MV, Fried I (2010) A sense of direction in human entorhinal cortex. *Proceedings of the National Academy of Sciences of the United States of America* 107:6487-6492.
- Jordan K, Schadow J, Wuestenberg T, Heinze HJ, Jancke L (2004) Different cortical activations for subjects using allocentric or egocentric strategies in a virtual navigation task. *Neuroreport* 15:135-140.
- Katayama K, Takahashi N, Ogawara K, Hattori T (1999) Pure topographical disorientation due to right posterior cingulate lesion. *Cortex; a journal devoted to the study of the nervous system and behavior* 35:279-282.
- Kelley AE, Smith-Roe SL, Holahan MR (1997) Response-reinforcement learning is dependent on N-methyl-D-aspartate receptor activation in the nucleus accumbens core. *Proceedings of the National Academy of Sciences of the United States of America* 94:12174-12179.
- Kennerley SW, Walton ME, Behrens TE, Buckley MJ, Rushworth MF (2006) Optimal decision making and the anterior cingulate cortex. *Nat Neurosci* 9:940-947.
- Kim H, Shimojo S, O'Doherty JP (2006) Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS biology* 4:e233.

- Klein-Flugge MC, Kennerley SW (2016) Neural Signatures of Value Comparison in Human Cingulate Cortex during Decisions Requiring an Effort-Reward Trade-off. *36:10002-10015*.
- Knutson B, Cooper JC (2005) Functional magnetic resonance imaging of reward prediction. *Current opinion in neurology 18:411-417*.
- Konorski J (1967) Integrative activity of the brain. Chicago: University of Chicago Press.
- Laczo J, Andel R, Vyhnalek M, Vlcek K, Magerova H, Varjassyova A, Tolar M, Hort J (2010) Human analogue of the morris water maze for testing subjects at risk of Alzheimer's disease. *Neuro-degenerative diseases 7:148-152*.
- Lambrey S, Doeller C, Berthoz A, Burgess N (2012) Imagining being somewhere else: neural basis of changing perspective in space. *Cerebral cortex (New York, NY : 1991) 22:166-174*.
- Latini-Corazzini L, Nesa MP, Ceccaldi M, Guedj E, Thinus-Blanc C, Cauda F, Dagata F, Peruch P (2010) Route and survey processing of topographical memory during navigation. *Psychol Res 74:545-559*.
- Lee D, Seo H, Jung MW (2012) Neural basis of reinforcement learning and decision making. *Annu Rev Neurosci 35:287-308*.
- Logothetis NK (2008) What we can do and what we cannot do with fMRI. *Nature 453:869-878*.
- Lynch K (1960) *The Image of The City*. Cambridge: MIT Press.
- Maguire EA, Frackowiak RS, Frith CD (1997) Recalling routes around london: activation of the right hippocampus in taxi drivers. *The Journal of neuroscience : the official journal of the Society for Neuroscience 17:7103-7110*.
- Maguire EA, Burgess N, O'Keefe J (1999) Human spatial navigation: cognitive maps, sexual dimorphism, and neural substrates. *Current opinion in neurobiology 9:171-177*.
- Maguire EA, Burgess N, Donnett JG, Frackowiak RS, Frith CD, O'Keefe J (1998) Knowing where and getting there: a human navigation network. *Science (New York, NY) 280:921-924*.

- Maguire EA, Gadian DG, Johnsrude IS, Good CD, Ashburner J, Frackowiak RS, Frith CD (2000) Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences of the United States of America* 97:4398-4403.
- Marchette SA, Bakker A, Shelton AL (2011) Cognitive mappers to creatures of habit: differential engagement of place and response learning mechanisms predicts human navigational behavior. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31:15264-15268.
- Marchette SA, Vass LK, Ryan J, Epstein RA (2014) Anchoring the neural compass: coding of local spatial reference frames in human medial parietal lobe. *Nat Neurosci* 17:1598-1606.
- Mata R, von Helversen B, Rieskamp J (2010) Learning to choose: Cognitive aging and strategy selection learning in decision making. *Psychology and aging* 25:299-309.
- Matsumoto M, Matsumoto K, Abe H, Tanaka K (2007) Medial prefrontal cell activity signaling prediction errors of action values. *Nat Neurosci* 10:647-656.
- McCloskey M (2001) Spatial representation in mind and brain. In: *the handbook of cognitive neuropsychology: What deficits reveal about the human mind* (Rapp B, ed), pp 101-132. Philadelphia: Psychology Press.
- Miller AM, Vedder LC, Law LM, Smith DM (2014) Cues, context, and long-term memory: the role of the retrosplenial cortex in spatial cognition. *Frontiers in human neuroscience* 8:586.
- Moffat SD, Resnick SM (2002) Effects of age on virtual environment place navigation and allocentric cognitive mapping. *Behavioral neuroscience* 116:851-859.
- Moffat SD, Elkins W, Resnick SM (2006) Age differences in the neural systems supporting human allocentric spatial navigation. *Neurobiology of aging* 27:965-972.
- Montello DR (1991) The measurement of cognitive distance: Methods and construct validity. *Journal of Environmental Psychology* 11:21.
- Morganti F, Riva G (2014) Virtual reality as allocentric/egocentric technology for the assessment of cognitive decline in the elderly. *Studies in health technology and informatics* 196:278-284.

- Morris RG, Garrud P, Rawlins JN, O'Keefe J (1982) Place navigation impaired in rats with hippocampal lesions. *Nature* 297:681-683.
- Nadel L (2012) Cognitive Maps. In: *Handbook of Spatial Cognition* (Waller D, Nadel L, eds). Washington: American Psychological Association.
- Newhouse P, Newhouse C, Astur RS (2007) Sex differences in visual-spatial learning using a virtual water maze in pre-pubertal children. *Behavioural brain research* 183:1-7.
- O'Doherty J, Winston J, Critchley H, Perrett D, Burt DM, Dolan RJ (2003a) Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia* 41:147-155.
- O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* (New York, NY) 304:452-454.
- O'Doherty JP, Cockburn J, Pauli WM (2017) Learning, Reward, and Decision Making. *Annual review of psychology* 68:73-100.
- O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ (2003b) Temporal difference models and reward-related learning in the human brain. *Neuron* 38:329-337.
- O'Keefe J, Nadel L (1978) *The Hippocampus as A Cognitive Map*. Oxford: Clarendon Press.
- Ongur D, Price JL (2000) The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cerebral cortex* (New York, NY : 1991) 10:206-219.
- Osawa A, Maeshima S, Kunishio K (2008) Topographic disorientation and amnesia due to cerebral hemorrhage in the left retrosplenial region. *European neurology* 59:79-82.
- Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. *Nature* 441:223-226.
- Park DC, Polk TA, Mikels JA, Taylor SF, Marshuetz C (2001) Cerebral aging: integration of brain and behavioral models of cognitive function. *Dialogues in clinical neuroscience* 3:151-165.

- Park S, Chun MM (2009) Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *NeuroImage* 47:1747-1756.
- Parslow DM, Rose D, Brooks B, Fleming S, Gray JA, Giampietro V, Brammer MJ, Williams S, Gasston D, Andrew C, Vythelingum GN, Loannou G, Simmons A, Morris RG (2004) Allocentric spatial memory activation of the hippocampal formation measured with fMRI. *Neuropsychology* 18:450-461.
- Pengas G, Hodges JR, Watson P, Nestor PJ (2010) Focal posterior cingulate atrophy in incipient Alzheimer's disease. *Neurobiology of aging* 31:25-33.
- Pengas G, Williams GB, Acosta-Cabronero J, Ash TW, Hong YT, Izquierdo-Garcia D, Fryer TD, Hodges JR, Nestor PJ (2012) The relationship of topographical memory performance to regional neurodegeneration in Alzheimer's disease. *Frontiers in aging neuroscience* 4:17.
- Peterson B, Wells M, Furness TA, Hunt E (1998) The Effects of the Interface on Navigation in Virtual Environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 42:1496-1500.
- Petrides M, Pandya DN (1999) Dorsolateral prefrontal cortex: comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. *The European journal of neuroscience* 11:1011-1036.
- Plassmann H, O'Doherty J, Rangel A (2007) Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27:9984-9988.
- Preuschoff K, Bossaerts P, Quartz SR (2006) Neural differentiation of expected reward and risk in human subcortical structures. *Neuron* 51:381-390.
- Preuschoff K, Quartz SR, Bossaerts P (2008) Human insula activation reflects risk prediction errors as well as risk. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28:2745-2752.
- Quilodran R, Rothe M, Procyk E (2008) Behavioral shifts and action valuation in the anterior cingulate cortex. *Neuron* 57:314-325.
- Rangel A, Camerer C, Montague PR (2008) A framework for studying the neurobiology of value-based decision making. *Nature reviews Neuroscience* 9:545-556.



- Richardson AE, Montello DR, Hegarty M (1999) Spatial knowledge acquisition from maps and from navigation in real and virtual environments. *Memory & cognition* 27:741-750.
- Robertson CE, Hermann KL, Mynick A, Kravitz DJ, Kanwisher N (2016) Neural Representations Integrate the Current Field of View with the Remembered 360 degrees Panorama in Scene-Selective Cortex. *Current biology : CB* 26:2463-2468.
- Robinson JL, Barron DS, Kirby LA, Bottenhorn KL, Hill AC, Murphy JE, Katz JS, Salibi N, Eickhoff SB, Fox PT (2015) Neurofunctional topography of the human hippocampus. *Human brain mapping* 36:5018-5037.
- Rodgers MK, Sindone JA, 3rd, Moffat SD (2012) Effects of age on navigation strategy. *Neurobiology of aging* 33:202.e215-222.
- Rolls ET, O'Mara SM (1995) View-responsive neurons in the primate hippocampal complex. *Hippocampus* 5:409-424.
- Rolls ET, Kringelbach ML, de Araujo IE (2003) Different representations of pleasant and unpleasant odours in the human brain. *The European journal of neuroscience* 18:695-703.
- Roulet P, Sargolini F, Oliverio A, Mele A (2001) NMDA and AMPA antagonist infusions into the ventral striatum impair different steps of spatial information processing in a nonassociative task in mice. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 21:2143-2149.
- Ruddle RA, Payne SJ, Jones DM (1997) Navigating buildings in "desk-top" virtual environments: Experimental investigations using extended navigational experience. *Journal of Experimental Psychology: Applied* 3:143-159.
- Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. *Science (New York, NY)* 310:1337-1340.
- Samanez-Larkin GR, Kuhnen CM, Yoo DJ, Knutson B (2010) Variability in nucleus accumbens activity mediates age-related suboptimal financial risk taking. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30:1426-1434.
- Sargolini F, Florian C, Oliverio A, Mele A, Roulet P (2003) Differential involvement of NMDA and AMPA receptors within the nucleus accumbens in consolidation of

- information necessary for place navigation and guidance strategy of mice. *Learning & memory* (Cold Spring Harbor, NY) 10:285-292.
- Sato N, Sakata H, Tanaka YL, Taira M (2006) Navigation-associated medial parietal neurons in monkeys. *Proceedings of the National Academy of Sciences of the United States of America* 103:17001-17006.
- Schinazi VR, Nardi D, Newcombe NS, Shipley TF, Epstein RA (2013) Hippocampal size predicts rapid learning of a cognitive map in humans. *Hippocampus* 23:515-528.
- Schmelter A, Jansen P, Heil M (2009) Empirical evaluation of virtual environment technology as an experimental tool in developmental spatial cognition research. *European Journal of Cognitive Psychology* 21:724-739.
- Schroeder JP, Wingard JC, Packard MG (2002) Post-training reversible inactivation of hippocampus reveals interference between memory systems. *Hippocampus* 12:280-284.
- Schultheis MT, Himmelstein J, Rizzo AA (2002) Virtual reality and neuropsychology: upgrading the current tools. *The Journal of head trauma rehabilitation* 17:378-394.
- Schultz W (1998) Predictive reward signal of dopamine neurons. *Journal of neurophysiology* 80:1-27.
- Serino S, Riva G (2013) Getting lost in Alzheimer's disease: a break in the mental frame syncing. *Medical hypotheses* 80:416-421.
- Seymour B, O'Doherty JP, Dayan P, Koltzenburg M, Jones AK, Dolan RJ, Friston KJ, Frackowiak RS (2004) Temporal difference models describe higher-order learning in humans. *Nature* 429:664-667.
- Sharp T, Zetterstrom T, Ungerstedt U (1986) An in vivo study of dopamine release and metabolism in rat brain regions using intracerebral dialysis. *Journal of neurochemistry* 47:113-122.
- Shipman SL, Astur RS (2008) Factors affecting the hippocampal BOLD response during spatial memory. *Behavioural brain research* 187:433-441.

- Siegel AW, White SH (1975) The development of spatial representations of large-scale environments. In: *Advances in child development and behaviour*, pp 9-55. New York: Academic Press.
- Simon DA, Daw ND (2011) Neural correlates of forward planning in a spatial decision task in humans. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31:5526-5539.
- Sitzmann V, Serrano A, Pavel A, Agrawala M, Gutierrez D, Masia B, Wetzstein G (2018) Saliency in VR: How Do People Explore Virtual Environments? *IEEE transactions on visualization and computer graphics* 24:1633-1642.
- Small DM, Gregory MD, Mak YE, Gitelman D, Mesulam MM, Parrish T (2003) Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron* 39:701-711.
- Smith DM, Mizumori SJ (2006) Hippocampal place cells, context, and episodic memory. *Hippocampus* 16:716-729.
- Spiers HJ, Maguire EA (2006) Thoughts, behaviour, and brain dynamics during navigation in the real world. *NeuroImage* 31:1826-1840.
- Spiers HJ, Burgess N, Hartley T, Vargha-Khadem F, O'Keefe J (2001) Bilateral hippocampal pathology impairs topographical and episodic memory but not visual pattern matching. *Hippocampus* 11:715-725.
- Squire LR, Stark CEL, Clark RE (2004) THE MEDIAL TEMPORAL LOBE. *Annual Review of Neuroscience* 27:279-306.
- Stalnaker TA, Calhoun GG, Ogawa M, Roesch MR, Schoenbaum G (2010) Neural correlates of stimulus-response and response-outcome associations in dorsolateral versus dorsomedial striatum. *Frontiers in integrative neuroscience* 4:12.
- Stevens FL, Hurley RA, Taber KH (2011) Anterior cingulate cortex: unique role in cognition and emotion. *The Journal of neuropsychiatry and clinical neurosciences* 23:121-125.
- Sutherland RJ, Rodriguez AJ (1989) The role of the fornix/fimbria and some related subcortical structures in place learning and memory. *Behavioural brain research* 32:265-277.

- Sutton RS, Barto AG (1998) Eligibility Traces. In: Reinforcement Learning: An Introduction. Cambridge: MIT Press.
- Takahashi K, Meilinger T, Watanabe K, Bulthoff HH (2013) Psychological influences on distance estimation in a virtual reality environment. *Frontiers in human neuroscience* 7:580.
- Takahashi N, Kawamura M (2002) Pure topographical disorientation--the anatomical basis of landmark agnosia. *Cortex; a journal devoted to the study of the nervous system and behavior* 38:717-725.
- Tanaka SC, Doya K, Okada G, Ueda K, Okamoto Y, Yamawaki S (2004) Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nat Neurosci* 7:887-893.
- Taube JS (2007) The head direction signal: origins and sensory-motor integration. *Annu Rev Neurosci* 30:181-207.
- Taylor HA, Brunye TT (2012) I Go Right, North, and Over: Processing Spatial Language. In: *Handbook of Spatial Cognition* (Nadel L, Waller D, eds). Washington: American Psychological Association.
- Thorndyke PW, Hayes-Roth B (1982a) Differences in spatial knowledge acquired from maps and navigation. *Cognitive psychology* 14:560-589.
- Thorndyke PW, Hayes-Roth B (1982b) Differences in spatial knowledge acquired from maps and navigation. *Cognitive psychology* 14:560-589.
- Tolman EC (1948) Cognitive map in rats and men. *The Psychological Review* 55:189-208.
- Tom SM, Fox CR, Trepel C, Poldrack RA (2007) The neural basis of loss aversion in decision-making under risk. *Science (New York, NY)* 315:515-518.
- Tremblay L, Schultz W (1999) Relative reward preference in primate orbitofrontal cortex. *Nature* 398:704-708.
- Tricomi E, Balleine BW, O'Doherty JP (2009) A specific role for posterior dorsolateral striatum in human habit learning. *The European journal of neuroscience* 29:2225-2232.

- Valentin VV, Dickinson A, O'Doherty JP (2007) Determining the neural substrates of goal-directed learning in the human brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27:4019-4026.
- van der Ham IJ, Faber AM, Venselaar M, van Kreveld MJ, Loffler M (2015) Ecological validity of virtual environments to assess human navigation ability. *Frontiers in psychology* 6:637.
- Van Hoesen GW, Pandya DN (1975) Some connections of the entorhinal (area 28) and perirhinal (area 35) cortices of the rhesus monkey. III. Efferent connections. *Brain research* 95:39-59.
- Vann SD, Aggleton JP, Maguire EA (2009) What does the retrosplenial cortex do? *Nature reviews Neuroscience* 10:792-802.
- Viard A, Doeller CF, Hartley T, Bird CM, Burgess N (2011) Anterior hippocampus and goal-directed spatial decision making. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31:4613-4621.
- Waller D, Loomis JM, Steck SD (2003) Inertial cues do not enhance knowledge of environmental layout. *Psychonomic bulletin & review* 10:987-993.
- Waller D, Beall AC, Loomis JM (2004) Using virtual environments to assess directional knowledge. *Journal of Environmental Psychology* 24:105-116.
- Wallis JD, Miller EK (2003) Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *The European journal of neuroscience* 18:2069-2081.
- Walton ME, Devlin JT, Rushworth MF (2004) Interactions between decision making and performance monitoring within prefrontal cortex. *Nat Neurosci* 7:1259-1265.
- Wiener JM, Büchner SJ, Hölscher C (2009) Taxonomy of Human Wayfinding Tasks: A Knowledge-Based Approach. *Spatial Cognition and Computation* 9:13.
- Wiener JM, Kmecova H, de Condappa O (2012) Route repetition and route retracing: effects of cognitive aging. *Frontiers in aging neuroscience* 4:7.
- Wiener JM, de Condappa O, Harris MA, Wolbers T (2013) Maladaptive bias for extrahippocampal navigation strategies in aging humans. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33:6012-6017.

- Wilson CJ, Soranzo A (2015) The Use of Virtual Reality in Psychology: A Case Study in Visual Perception. *Computational and mathematical methods in medicine* 2015:151702.
- Wilson CR, Gaffan D, Browning PG, Baxter MG (2010) Functional localization within the prefrontal cortex: missing the forest for the trees? *Trends in neurosciences* 33:533-540.
- Wilson MA, McNaughton BL (1993) Dynamics of the hippocampal ensemble code for space. *Science (New York, NY)* 261:1055-1058.
- Wingard JC, Packard MG (2008) The amygdala and emotional modulation of competition between cognitive and habit memory. *Behavioural brain research* 193:126-131.
- Witmer BG, Bailey JH, Knerr BW, Parsons KC (1996) Virtual spaces and real world places: transfer of route knowledge. *International Journal of Human-Computer Studies* 45:413-428.
- Wolbers T, Büchel C (2005) Dissociable retrosplenial and hippocampal contributions to successful formation of survey representations. *Journal of Neuroscience* 25:3333-3340.
- Wolbers T, Wiener JM (2014) Challenges for identifying the neural mechanisms that support spatial navigation: the impact of spatial scale. *Frontiers in human neuroscience* 8.
- Wolbers T, Weiller C, Büchel C (2004) Neural foundations of emerging route knowledge in complex spatial environments. *Cognitive Brain Research* 21:401-411.
- Wolbers T, Hegarty, M. (2010) What determines our navigational abilities? *Trends in Cognitive Sciences* 14:138-146.
- Worthy DA, Gorlick MA, Pacheco JL, Schnyer DM, Maddox WT (2011) With age comes wisdom: decision making in younger and older adults. *Psychological science* 22:1375-1380.
- Wunderlich K, Rangel A, O'Doherty JP (2009) Neural computations underlying action-based decision making in the human brain. *Proceedings of the National Academy of Sciences of the United States of America* 106:17199-17204.

- Wunderlich K, Dayan P, Dolan RJ (2011) Mapping value based planning and extensively trained choice in the human brain. *Nature Neuroscience* 15:786 - 791.
- Wunderlich K, Smittenaar P, Dolan RJ (2012) Dopamine enhances model-based over model-free choice behavior. *Neuron* 75:418-424.
- Yin HH, Knowlton BJ (2004) Contributions of striatal subregions to place and response learning. *Learning & memory (Cold Spring Harbor, NY)* 11:459-463.
- Zaehle T, Jordan K, Wustenberg T, Baudewig J, Dechent P, Mast FW (2007) The neural basis of the egocentric and allocentric spatial frame of reference. *Brain research* 1137:92-103.





# List of Abstract and Publications

## Submitted Manuscript (under review)

9<sup>th</sup> April 2018

Scientific Reports

*Neural Signature of Reinforcement Learning Correlates with Strategy Adoption during Spatial Navigation*

Authors: D. Anggraini, S. Glasauer, K. Wunderlich

## Manuscript in Preparation

*Of Reward, Age, and Space: Reinforcement Learning Algorithms Explain Age-related Changes in Spatial Navigation*

Authors: D. Anggraini, S. Glasauer, V. Flanagan, O. Prat, K. Wunderlich

## Conference Posters

9<sup>th</sup> – 11<sup>th</sup> July 2015

IMPRS NeuroCom, Leipzig, Germany

*To Locate The Rewards: Can Reinforcement Learning Help One from Getting Lost?*

Authors: D. Anggraini, S. Glasauer, K. Wunderlich

25<sup>th</sup> – 27<sup>th</sup> September 2015

SNE Annual Meeting, Florida, USA

*To Locate My Rewards: Can Reinforcement Learning Help Me from Getting Lost?*

Authors: D. Anggraini, S. Glasauer, K. Wunderlich

26<sup>th</sup> – 30<sup>th</sup> June 2016

iNAV 2016, Bad Gastein, Austria

*Route-based Learning and Map-based Planning during Navigation: Contributions of the Reward System*

Authors: D. Anggraini, S. Glasauer, K. Wunderlich

2<sup>nd</sup> – 6<sup>th</sup> July 2016 10th FENS Forum of Neuroscience, Denmark

*Neural Correlates of Value-based Choices during Spatial Navigation*

Authors: D. Anggraini, S. Glasauer, K. Wunderlich

28<sup>th</sup> – 30<sup>th</sup> August 2016 SNE Annual Meeting, Berlin, Germany

*Route-based Learning and Map-based Planning during Navigation: Contributions of the Reward System*

Authors: D. Anggraini, S. Glasauer, K. Wunderlich

### **Symposium/Conference Talk**

29<sup>th</sup> July – 1<sup>st</sup> August 2015 NeuroDoWo 2015, Cologne, Germany

Reinforcement Learning Model of Spatial Decision Making

Authors: D. Anggraini, S. Glasauer, K. Wunderlich

6<sup>th</sup> – 8<sup>th</sup> October 2017 SNE Annual Meeting, Toronto, Canada

Gratitude and Pride: Neural Correlates of Reward Contribution to Others and Onself

Authors: K. Ding, D. Anggraini, K. Wunderlich

### **Science Communication Talk**

11<sup>th</sup> – 14<sup>th</sup> May 2017 I, Scientist Conference, Berlin, Germany

*Science Slam: Reinforcement Learning, Navigation, and the Brain*

21<sup>st</sup> July 2017 15x4 Munich, Munich, Germany

*Neuroeconomics: Because Brand Matters*

16<sup>th</sup> November 2017 15x4 Munich, Munich, Germany

*Cognitive Reserve: How the Brain Deals with Changes*

# Eidesstattliche Versicherung Affidavit

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation:

„Decision in Space: Reinforcement Learning Models Predict Individual Differences in and Neural Substrates of Strategy Adoption in Spatial Navigation“

selbständig angefertigt habe, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

I hereby confirm that the dissertation “Decision in Space: Reinforcement Learning Models Predict Individual Differences and Neural Substrates of Strategy Adoption in Spatial Navigation” is the result of my own work and that I have only used sources or materials listed and specified in the dissertation.

Munich, 15. May 2018

ir. Dian Anggraini, M.Sc.



# List of Author Contributions

**Manuscript 1: Neural Signatures of Reinforcement Learning Correlates with Strategy Adoption in Spatial Navigation\*** First Author: D.A.; Conceptualization: D.A., S.G., and K.W.; Methodology: D.A, S.G., and K.W.; Software: D.A. and K.W.; Validation: D.A, S.G., and K.W.; Resources: S.G.; Investigation: D.A.; Formal Analysis: D.A. and K.W.; Writing – Original Draft: D.A. and K.W.; Writing – Review and Editing: D.A., S.G., and K.W.; Visualization: D.A. and K.W.

**Manuscript 2: Of Reward, Age, and Space: Reinforcement Learning Algorithms Explain Age-related Changes in Spatial Navigation\*** First Author: D.A.; Conceptualization: D.A., S.G., V.F., and K.W.; Methodology: D.A, S.G., V.F., and K.W.; Software: D.A. and K.W.; Validation: D.A, S.G., and K.W.; Resources: S.G.; Investigation: D.A. and O.P.; Formal Analysis: D.A., O.P., and K.W.; Writing – Original Draft: D.A. and O.P.; Writing – Review and Editing: D.A., S.G., V.F., and K.W.; Visualization: D.A. and K.W.

\*D.A: Dian Angraini; S.G: Stefan Glasauer; V.F: Virginia Flanagin; O.P: Ot Prat; K.W: Klaus Wunderlich.

I hereby confirm the specified contributions for the preparation of the manuscripts.

Munich, 15 May 2018

ir. Dian Anggaini, M.Sc.

Dr. Klaus Wunderlich