NAIST-IS-DD0861212

Doctoral Dissertation

Multiple Reinforcement Learning Action Selection Strategies in Prefrontal-Basal Ganglia and Cerebellar Networks

Alan de Souza Rodrigues

September 14, 2011

Division of Applied Informatics
Graduate School of Information Science
Nara Institute of Science and Technology

A Doctoral Dissertation
submitted to the Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of SCIENCE

Alan de Souza Rodrigues

Thesis Committee:

Professor Kazushi Ikeda	(Supervisor)
Professor Kenji Doya	(Co-supervisor)
Professor Kotaro Minato	(Co-supervisor)
Associate Professor Tomohiro Shibata	(Co-supervisor)
Associate Professor Junichiro Yoshimoto	(Co-supervisor)

Multiple Reinforcement Learning Action Selection Strategies in Prefrontal-Basal Ganglia and Cerebellar Networks*

Alan de Souza Rodrigues

Abstract

Humans can learn to select actions from scratch, by trial-and-error, or by using knowledge from previous experiences. Also, a thoughtful deliberative strategy can be used for planning of future behaviors, whereas a reactive strategy for selection of learned sensory-motor mappings. Reinforcement Learning (RL), a computational theory of adaptive optimal control, proposes two formal methods for action selection: Model-Free (MF) uses action values to predict future rewards based on current states and available actions and Model-Based (MB) uses a forward model by mental simulation to predict future states reached by hypothetical actions. It has been suggested that humans might use MF and MB strategies which are implemented locally by distinct brain areas or in neural networks linking the prefrontal cortex, basal-ganglia and cerebellum. The goal of this thesis is to investigate whether and how humans use multiple RL action selection strategies, and where in the brain they are implemented.

We hypothesized based on RL a parallel model for action selection and learning from reward and trial-and-error. When knowledge of the task dynamics is unavailable, a MF valued-based method is used for exploration and learning predictions of future rewards from sensory states and actions. If an internal model exists, a MB forward method is implemented by mental simulation to predict future states reached by hypothetical actions. After repeated experiences, a MF memory-based method is employed for fast selection of actions found successful for a given sequence of sensory states.

This hypothesis was experimentally tested as human subjects performed a ‘grid-sailing task’ and learned to move a cursor from a start position to a target goal with the shortest finger movement sequence by button pressing. Performance feedback was provided as a reward score at the trial end. Subjects extensively learned action sequences for fixed

key-mappings (KM) and start-goal (SG) sets in a training session and were later tested under three task conditions: a) Condition 1, new KM, b) Condition 2, learned KM, and c) Condition 3, learned KM-SG sets, practiced in the training session. The response start time was manipulated by introducing a delay period of about 4~6s preceding the response start go signal in half of the trials.

Behavior analysis revealed distinct performance profiles in the test session: a) exploratory, variable and slow learning in Condition 1 due to the use of the new KM; b) fast learning and high reward score especially in trials with the delay period in Condition 2 as subjects could use the learned KM, c) accurate performance with fast sequence execution in Condition 3 which required the learned KM-SG sets in the training session. These results suggest that a value-based exploratory method was used in Condition 1; a model-based planning method in Condition 2 and the memory-based habitual method in Condition 3.

The fMRI analysis of the brain signals during the delay period preceding movement execution revealed predominant activity by the use of: a) value-based method in Condition 1 in the medial orbital frontal cortex, ventromedial striatum and left cerebellum; b) model-based method in Condition 2 in the dorsolateral prefrontal cortex, dorsomedial striatum and right cerebellum, and c) memory-based method in Condition 3 in the supplementary motor area, right dorsolateral striatum and right anterior cerebellum.

We could demonstrate by the behavior analysis that humans use multiple action selection strategies depending on their experiences with a given task, existence of an accurate internal model and available time for thinking and use of the model. The fMRI results showed that an action selection method is implemented in the brain by interaction of multiple brain areas, such as networks linking the prefrontal cortex, the basal-ganglia and cerebellum, and not locally as had been previously speculated. What computational operations these brain regions play in mental simulation, planning and preparation of future behaviors and how they are disrupted in psychiatric conditions is still a problem to be elucidated.

Keywords:
reinforcement learning, model-free, model-based, action selection, sequence learning, prefrontal cortex, basal ganglia, cerebellum

.*Doctoral Dissertation, Division of Applied Informatics, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD0861212, September 14, 2011.

強化学習理論に基づく 意思決定戦略における 前頭前野ー大脳基底核ー小脳系神経回路の計算論的機能に関する 研究

アレン　デ　ソウザ　ホドリゲス

Abstract

ヒトは，試行錯誤で，または過去の経験に基づいて，行動を一から決定・学習することができる．また，ヒトは，既知の感覚ー運動変換に基づいて反射的な行動戦略をとる一方で，熟考した複雑な戦略で，将来の行動を計画することができる．適応制御のための学習理論である強化学習 (Reinforcement Learning: RL) は，ふたつの行動戦略を提案している．一方はモデルフリー (Model-Free: MF) で，将来の報酬を現在の状態における各行動の価値から予測する．他方はモデルベース (Model-Based: MB) で，例えばメンタルシミュレーションによって，行動に基づく状態遷移を順モデルから推定する．これまで，ヒトは，MF・MB両方の戦略を用いることが出来ると提案されてきた．また，MF・MB戦略は，脳の異なる領域に局所的に実装されていると考えられてきた． 本論文では，この並列的な強化学習戦略を，ヒトがどのように用いているのかを調べる．また，各戦略が，脳のどこで実装されているかを調べる．

本研究で私達は，以下のように，行動選択には，複数の強化学習戦略が並列的・段階的に関与すると仮定した．まず，課題に対する知識が不足しているとき，ヒトは，MF戦略の価値依存的戦略 (Value-based method) で，状態・行動から報酬を予測し，探索的に行動を決定する．次に，課題の内部モデルがあるとき，ヒトは，メンタルシミュレーションで，行動に基づく状態遷移を順モデルで推定する．さらに，同一の課題を繰り返し経験した後では，ヒトは，MF戦略の記憶依存的戦略 (Memory-based method) で，状態に適した行動を，すばやく決定する．

私達は，この仮説を，格子探索課題 (Grid-sailing task) 遂行中のヒト被験者の行動から実験的に検証した．同課題で，被験者は，指でボタンを押し，コンピュータ上のカーソルを動かした．被験者は，格子世界で，スタート位置からゴール位置までを最短でむすぶ指運動系列を学習した．各試行の最後に，被験者の行動，すなわち，指運動系列のフィードバックとして，報酬を提示した．訓練セッションで，被験者は，指運動 - カーソル移動関係

(key-mappings: KM) とスタート - ゴール位置 (Start-goal set: SG) を，特定の組み合わせで学習した．訓練セッション後，被験者の行動を，以下の3つのテスト条件で調べた：a) 条件1，新規KM－新規SG；b) 条件2，学習済みKM－新規SG；c) 条件3，学習済みKM－学習済みSG．全試行中半数の試行では，指運動開始を指示する視覚刺激の前に，4~6秒の遅延時間を設定した．遅延時間ありとなしの試行で，被験者の反応時間を計測・比較した．

被験者の行動を解析した結果，3つのテスト条件で異なる行動が見られた：a) 新規KMを用いた条件1では，探索に基づくゆるやかな学習が見られた．b) 学習済みKMを用いた条件2，特に遅延時間ありの試行では，すばやい学習と高い報酬量が見られた．c) 学習済みのKM-SGを直接用いた条件3では，正確な行動とすばやい指運動系列が見られた．これらの結果は，条件1，条件2，条件3でそれぞれ，被験者がValue-based method，Model-based planning method，Memory-based habituation method を用いていたことを示唆する．

指運動開始前の遅延時間で，課題中の脳活動を機能的磁気共鳴画像法 (functional Magnetic Resonance Imaging: fMRI) で測定した結果，a) 条件1のValue-based methodでは，内側眼窩前頭皮質，腹内側線条体，左小脳，b) 条件2の Model-based methodでは，背外側前頭前野，背内側線条体，右小脳，c) 条件3のMemory-based methodでは，補足運動野，右背外側線条体，右前小脳が，それぞれ顕著に活動したことがわかった．

行動解析の結果，ヒトは，課題依存的に複数の行動戦略，すなわち，Value-based method，Model-based planning method，Memory-based habituation methodを使い分けることを，本研究は示唆した．また，fMRIの解析結果，これらの行動戦略は，前頭前野－大脳基底核－小脳系神経回路を含んだ複数の脳領野で実装されていることを示唆した．従来，行動戦略は，脳の各領野に局所的に実装されていると考えられており，本研究の結果とは異なる．今後，本研究で同定した神経回路網が，行動予測の準備・予測・メンタルシミュレーションのどの段階・計算理論に関与するのか，また，精神病でどのように阻害されるのかを検証する．

キーワード:
強化学習，モデルフリー戦略，モデルベース戦略，行動選択， 系列学習，前頭前野，大脳基底核，小脳

.*奈良先端科学技術大学院大学 情報科学研究科 システム情報学領域 博士論文, NAIST-IS-DD0861212, 2011年9月14日.

ACKNOWLEDGEMENTS

I am truly indebted, sincerely and heartily grateful to my supervisor Kenji Doya. First and foremost for his decision to accept a naive boy from the Amazon as his student. It certainly was a risk decision and I still wonder if he will ever profit from it. I myself have, however, learned so much from his computational and neuroscience knowledge, the importance of defining new scientific problems, as well as clarity in scientific writing and social networking. This dissertation would not have been possible without his patience, support and guidance.

I also would like to show my gratitude to my formal supervisors, first Shin Ishii who followed me throughout my Master's degree and the first year as PhD student, and my current supervisor Kazushi Ikeda. Tomohiro Shibata has also been essential in setting up the way for my coming to Japan as well as my academic life. Junichiro Yoshimoto has also played an important role in the conduction of the behavioral and fMRI experiments, data analysis and his support in my academic life. I thankful to Minato Kotaro for being a member in my thesis committee. Makoto Ito has also been invaluable with his support in data analysis as well as with continuous discussions and theoretical interpretation of the behavioral and fMRI results.

I am very honored to have been given the chance to meet and discuss about my research with so many famous researchers including Kazuyuki Samejima, Paul Cisek, John O'Doherty, Okihide Hikosaka, Leo Cohen, Read Montague, Peter Dayan, Daniel Wolpert, Ethan Bromberg-Martin, Hiroshi Imamizu, Ann M. Graybiel, Nicolas Schweighofer, Kazuyuki Seki, Masamichi Sakagami, Xiao-Jing Wang, Reza Shadmehr, Francisco Valero-Cuevas, Gordon Arbuthnott, David Redish, Saori Tanaka and many others.

It was also a pleasure to befriend with the new generation of talented and promising young researchers with who I had the chance not only to discuss about research but also to have fun in conference meetings, summer schools, on the

beach, izakaya and many other places: Aki Funamizu, Makoto Otsuka, Viktor Zhumatiy, Shinichi Maeda, Kanemura Atsunori, Naoki Honda, Hayashi Kohei, Mauricio Burdelis, Akaishi Rei, Ben Huh, Yuval Tassa, Takashi Nakano, Tetsuro Morimura, Ken Kinjo, Naoto Yoshida, Romain Caze, Ping Wan, Ben Torben-Nielsen, Cathy Vickers, Yuki Hidaka, Jeong-Yoon Lee, Yuki Sakai, Mehdi Khamassi, Matt van der Meer, Cengiz Gunay, Atsushi Noritake, Chri Burke and many others.

A very special thanks to the members in the Neural Computation Unit: Katsuhiko Miyazaki, Kayoko Miyazaki, Eiji Uchibe, Stefan Elfwig, Makoto Ito, Ken Kinjo, Makoto Otsuka, Naoto Yoshida, Masato Hoshino, Viktor Zhumatiy, Takehiko Yoshida, Ryo Shiro, Tetsuro Morimoto, Takumi Kamioka, Takashi Nakano and Aki Funamizu. These people have become an extension of my family in Japan.

The work developed in this thesis has been possible with the diligent support of the Neural Computation Unit staff, including Emiko Asato, Tomofumi Inoue, Izumi Nagano, Hitomi Shinzato and Chikako Uehara. Emiko has been a big sister throughout these five years of graduate studies and helped me not only with administrative but also with support for my daily life needs.

I am very thankful to all those people who directly or indirectly supported my academic life at NAIST and OIST.

I am grateful to the Japanese Ministry of Education, Culture, Sports, Science and Technology for funding my Master's and PhD degrees.

I thank my parents for their support throughout my whole academic career and for enduring, together with my family, the long distance that separates us for six years now.

CONTENTS

LIST OF PUBLICATIONS

Alan Fermin (**Fermin, A.**) is the name I use for scientific publications.

Peer-Reviewed Papers

Fermin, A., Yoshida, T., Ito, M., Yoshimoto, J., Doya, K. Evidence for model-based action planning in a sequential finger movement task. Journal of Motor Behavior, 42, 6, 371-379, November, 2011.

Fermin, A., Ito, M., Yoshimoto, J., Doya, K. Value-based, model-based, and memory-based learning strategies in the cortico-basal ganglia and cerebellar networks (in preparation).

Conference Abstract Proceedings

Fermin, A., Yoshida, T., Ito, M., Yoshimoto, J., Doya, K. (2010). Neural mechanisms for model-free and model-based reinforcement strategies in humans performing a multi-step navigation task. Neuroscience Research, 68, e285-e286.

Fermin, A., Yoshida, T., Tanaka, S., Ito, M., Yoshimoto, J., Doya, K. (2009). Reinforcement Learning Strategies for Sequential Action Learning. Neuroscience Research, 65, S236.

Fermin, A., Yoshida, T., Tanaka, S., Ito, M., Yoshimoto, J., Doya, K. (2009). Model-Free and Model-Based Reinforcement Learning Strategies in the Acquisition of Sequential Actions. Proceedings of the 19th Annual Conference of the Society for the Neural Control of Movement, i67.

Posters

Fermin, A., Yoshida, T., Ito, M., Yoshimoto, J., Doya, K. Neural circuits for model-free and model-based action selection strategies in multi-step action learning. 41st NIPS International Symposium, Okazaki, Japan, December 16th, 2010.

Fermin, A., Yoshida, T., Ito, M., Yoshimoto, J., Doya, K. Neural mechanisms for model-free and model-based reinforcement strategies in humans performing a multi-step navigation task. The 33rd Annual Meeting of the Japan Neuroscience Society, Kobe, Japan, 2010.

Fermin, A., Yoshida, T., Tanaka, S., Ito, M., Yoshimoto, J., Doya, K. Reinforcement Learning Strategies for Sequential Action Learning. 32nd Annual Meeting of the Japan Neuroscience Society, Nagoya, Japan, 2009.

Fermin, A., Yoshida, T., Tanaka, S., Ito, M., Yoshimoto, J., Doya, K. Candidate Neural Networks for Implementing Model-Free and Model-Based Reinforcement Learning Strategies in the Selection of Sequential Actions. 10th Summer Workshop on the Mechanisms of Brain and Mind, Sapporo, Japan, 2009/08/09.

LIST OF ABBREVIATIONS

ANOVA	analysis of variance
BA	Brodmann area
BG	basal ganglia
CB	cerebellum
DA	dopamine
DLPFC	dorsolateral prefrontal cortex
DM	decision making
dmSTR	dorsomedial striatum
dlSTR	dorsolateral striatum
ET	execution time
FC	fixation cross
ITI	Inter-trial interval
KM	key-mapping
fMRI	functional magnetic resonance imaging
M1	primary motor cortex
MAD	mean absolute deviation
MF	model-free
MB	model-based
ML	motor learning
OAS	optimal action sequence
PC	parietal cortex
PFC	prefrontal cortex
RL	reinforcement learning
RT	reaction time
ROI	region of interest
SG	start-goal positions
SPM	statistical parametric mapping
NOAS	non-optimal action sequence
preSMA	pre-supplementary motor area

SMA	supplementary motor area
SNc	substantia nigra pars compacta
SNr	substantia nigra pars reticulata
SR	stimulus-response
STR	striatum
pre-SMA	pre-supplementary motor area
vmSTR	ventro-medial striatum
VTA	ventral tegmental area

LIST OF FIGURES

CHAPTER 1

Introduction

Humans and other vertebrates (mammals, birds, fishes) share the amazing capability, evolutionarily acquired, to continuously learn new motor and cognitive behaviors (Skinner, 1984; Skinner, 1975; Alcock, 2001; Bekoff, 2002). This ability, present from birth to advanced age, ceases only with death. Another common feature shared among different species is the existence of a central nervous system (Herculano-Houzel, 2009; Barton, 2006; Barton and Harvey, 2000; Roth and Dicke, 2005, Sultan, 2002) and/or of a peripheral nervous system that innervates their whole bodies (Murakami and Tanaka, 2011; Burish et al., 2010).

The human brain is highly complex and large when compared with the brains of non-human primates (Whiting et al, 2003; Ramnani, 2006) and it is estimated to contain about 100 billion (10^{11}) neurons (Pakkenberg, 1988, Herculano-Houzel and Lent, 2005). The brain of a bird is much smaller yet, it retains neural structures that resemble those of a human brain. The total number of neurons in a songbird brain is about 20 millions ($2x10^{7}$). The insect nervous system is restricted to only a few hundred or a few thousand

cells. The nematode worm Caenorhabditis elegans (C. elegans) has just 302 neurons (Felix, et al, 2010).

Regardless of the number of neurons in a nervous system, some species are able to learn new behaviors and to retrieve and execute learned ones. Therefore, these species possess common learning and memory mechanisms capable of identifying information arriving from the outside world, generate a response or a sequence of responses, store those responses if considered important and retrieve them for execution when encountering the same situation.

Learning, memory and retrieval, however, are not be the best evolutionary solutions for survival in a dynamically changing and uncertain world. Humans evolved and organized in social networks either by attachment/emotional bonds or by social rules and conventions. We became able to understand the intentions and desires of others in our surrounding and for that we developed complex communication skills by gestures, speech, writing, dancing and arts. Most importantly, humans became able to imagine and contemplate the future, to think ahead, make decisions, simulate and evaluate the consequences of one's own actions as well as the outcomes achieved by others' behaviors (Yoshida and Ishii, 2006; ; Gusnard et al, 2001; Bar, 2007; Addis et al, 2007).

In a world in constant change, where people make decisions but immediately change their minds, it would be almost impossible to live a good social life without the ability to make plans and imagine what is going on inside others' mind. How has the brain found a

solution to this problem? How were we, humans, able to add to our fundamental learning and memory abilities, one more process, that of planning, prediction and imagination?

The human brain got larger with evolution (Haug, 1987; Andersen et al, 1992). The brain region whose total area most prominently enlarged is the prefrontal cortex. The human prefrontal cortex is about two or three times the size of that of a chimpanzee and even several times larger than the brain of new world monkeys (Whiting and Barton, 2003). The lateral hemispheres of the cerebellum cerebellum paralleled a similar enlargement as that of PFC, in a much larger proportion (Lange, 1975), though. The cerebellum lateral lobules, Crus 1 and Crus 2, expanded dramatically and almost doubled the size of the cerebellar cortex (Ramnani, 2006, Llinas and Hilmman, 1969; Dow and Moruzzi, 1958). New anatomical connections were made with the enlargement of several brain areas. Reciprocal anatomical connections between the PFC, the basal ganglia and have been discovered only in the last 20 years with the development of novel neuronal fiber tracing methods (Alexander et al, 1986; Bostan and Strick, 2010).

Imaging experiments with humans and lesion and neurophysiological recording studies in monkeys have shown the involvement of PFC neurons in higher order cognitive processes such as planning sequences of motor actions and events (Shima et al, 2007; Ninokura, 2003), spatial memory tasks (Goldman-Rakic et al, 1990; Funahashi et al, 1991, Funahashi et al, 1992), problem-solving (Mushiake et al, 2009), planning final target goals (Saito et al, 2005, Mushiake et al, 2006), prediction of action outcomes (Glascher et al, 2010), mental simulation and imagining the future (Simon and Daw, 2011; Szpunar et al, 2007) and control of social behavior (Wood, 2006; Beer et al, 2006).

These results also point out that the basal ganglia and cerebellar hemispheres play roles in higher cognitive processes since these regions receive input from and send output back to the prefrontal cortex. Although the participation of the prefrontal cortex in cognitive and social processes has been well investigated, the role of basal ganglia and the cerebellum is still in under intense scrutiny.

Evolution, therefore, has shaped the brain with built-in functions for learning and memory. Social pressures and environmental demands requiring flexibility and quick adaptation of behavior led humans to acquire the capability for planning and simulation by using more elaborate cognitive processes. The results of neurophysiological and imaging experiments imply that the prefrontal cortex might be the neural substrate that allows humans to realize planning and prediction. This capacity for planning and simulation of future events might have been optimized with the enlargement of several brain areas as well as the anatomical connections, including those linking the prefrontal cortex, basal ganglia and cerebellum.

The basic processes of how humans learn new motor behaviors, store and retrieve motor memories or engage in complex planning processes for making action choices is well documented. Nevertheless, the actual strategies used for action selection under different contexts and levels of experience is still an open problem. Patients with neurological and psychiatric disorders, such as Parkinson's disease, cerebral palsy and developmental coordination disorder, are highly impaired in learning and planning motor behaviors (Muslimovic et al, 2007; Hutton et al, 1998; Sugden and Chambers, 2005), solving executive and memory tasks (Patniyot, 2011; Bralet et al, 2008), and engage in social

and communicative behaviors (Koegel and Koegel, 1990; Koegel et al, 1992; Carr and Durant, 1985; Kanner et al, 1972). Therefore, the clarification of the mechanisms used for action selection and their neural correlates is essential for the development of therapeutical interventions, such as for schizophrenia, Parkinson's disease and developmental disorders, including autism and cerebral palsy.

1.1. Motivation and Scope of this Thesis

The goal of this thesis is to understand whether and how humans use prediction models for action selection and learning, and what brain networks implement such strategies. It is known that the acquisition of new motor behaviors goes through a series of stage-wise transitions (Fitts and Posner, 1968; Ackerman, 1988; Adams, 1971) . Initially, long response time is needed for action selection and for that humans use higher-order cognitive and attentional processes to understand the task dynamics and to map which action to take to get which outcome. After some experiences and the formation of model of the environment, humans become able to use these learned representations for planning and simulating the consequences of candidate actions. Finally, after long-term experience and repeated practice, action selection becomes automatic and fast, and requires the less cognitive effort. This view implies that humans might utilize multiple action selection strategies in learning depending on their level of experience with a task.

Interestingly, the problem of action selection in decision making situations might parallel that of action selection in motor learning. Two main strategies for choosing actions are frequently reported in the decision making literature. A thoughtful, deliberative strategy is used for planning goal-directed behaviors by simulating state transitions and the consequences of hypothetical actions, whereas a reactive, habitual strategy is used for selection of well-learned sensory-motor mappings, such as in stimulus-response association learning. The deliberative and reactive decision making strategies correspond, to some extent, to the intermediate and advanced motor learning stages, respectively.

Reinforcement learning (RL, Sutton and Barto, 1998; Doya, 2007), a computational theory of adaptive optimal control, has attracted the attention of DM researchers and has proved powerful in providing normative descriptions of how humans (Tanaka, et al, 2005; Glascher et al, 2009; Wunderlich et al, 2009) and animals (Ito and Doya, 2009; Seo et al, 2009) select and learn actions. RL proposes two methods for action selection, model-free (MF) and model-based (MB), that resemble the reactive and deliberative strategies used by humans for motor learning and decision making.

The model-free method updates a policy using action value functions that predict future rewards based on current sensory input states and available actions by trial-and-error from actual exploratory experiences. In situations where the environment suddenly changes, the policy must be updated based on the difference of the predicted and actual delivered rewards. The model-based method employs an internal forward model, assuming that it has been previously acquired, to predict the future state reached by a hypothetical action or multi-step actions, and uses the results of this simulation to

evaluate states and actions, and to selecting an action for execution based on its goodness. A model-based method is helpful not only for on-line action planning, but also for off-line learning of a policy from limited experiences (Doya, 2007). Success of using a model-based method depends, however, on the model accuracy, working memory capacity, time available for computation and the action search depth and width.

Action selection in DM and ML might be guided by the same or, at least, common mental processes. The theory of RL provides formal descriptions of what selection strategies humans use in decision making and learning. However, it is still unclear whether and how humans use MF and MB RL strategies, the factors that influence and arbitrate the use of one strategy instead of the other, and what brain mechanisms are predominantly involved in implementing them.

We explore these different topics under different approaches in this thesis:

I) **Behavioral Studies**

Behavioral experiments were conducted to investigate whether and how humans use prediction models for action selection. Subjects were asked to play a computer game, the 'grid-sailing task', and perform sequential finger key-presses to move a cursor from its start position to a target goal. A cursor was associated with a three-motion direction key-mapping rule and subjects started off the task without knowing it. Subjects, therefore, had to learn it by trial-and-error as a reward score was provided as performance feedback at the trial end. A delay period of about 6 seconds preceded the go signal for action execution in half of the trials, whereas in the other half, responses should start

immediately after cue presentation. The task was designed to share an interface between ML and DM paradigms, where action selection could be implemented by strategies common to both fields. A training session was administered on day 1 and subjects extensively learned fixed sets of action sequences. After a one-night sleep subjects performed the main experiment under three task conditions which had the purpose to mimic and break apart the multiple action learning stages. Subjects' behavior data was analyzed under the approach of reinforcement learning (RL, Sutton and Barto, 1998; Doya, 2007), a computational theory of adaptive optimal control that is used to explain how humans learn actions by trial-and-error and available teaching signals, such as food rewards, money or verbal instructions.

II) **Brain Imaging Experiment**

A functional Magnetic Resonance Imaging (fMRI) experiment was carried out in order to investigate the brain networks activated as subjects performed the grid-sailing task under different experimental conditions. This analysis aimed at identifying the brain activity for the main trial events, namely: the delay period, the execution period and the reward score feedback period. This approach allowed us to separate the main cognitive processes that subjects engaged for planning and simulation of action sequences during the delay period from the motor-related processes involved in the action execution period, as well as the performance feedback-related processing during the reward period. Further analysis included the region of interest (ROI) method with a primary focus on the percent signal change in the activity of specific brain regions such as the PFC, BG and CB.

1.2. Outline of the Thesis

This thesis is organized as follows. In Chapter 2, a brief review of the literature describe the classical and contemporary theories of action selection and learning, the neural mechanisms involved in learning and open issues in the understanding of human motor behavior. Chapter 3 describes the preliminary behavioral results of an behavior experiment providing evidence that humans use a model-based action selection strategy for planning multistep actions. Based on these results we proposed a RL three-system parallel for action selection and learning that is able to explain why stage-wise transitions are observed in motor learning. Chapter 4 expands the results presented in Chapter 3 with a new task design and the results of fMRI data analysis. First, the experimental task design was modified in order to make it methodologically and statistically sound. The major change was the random assignment of subjects to one of three groups. And second, the task was performed inside the fMRI scanner allowing the recording and acquisition of brain signals during task execution. Finally, in Chapter 5 we summarize the main achievements and contributions presented in this thesis, and point out the future directions of this work.

CHAPTER 2

Action Learning: Behavior and Neural Control

2.1. Classical Theories of Motor learning

Humans and other animals have an extraordinary capability to learning new motor actions, to flexibly adjusting and adapting previously learned ones as well as retrieving old and recent motor memories imprinted in their brains. Nature and evolution endowed humans with built-in biological functions to learn and retrieve motor memories and other representations of the external world stored in the nervous system.

Learning functions come to work early in infancy and are expressed as newborn babies learn to control their bodies for reaching (Konzack and Dichgans, 1997; Rochat and Goubet, 1995; Thelen et al., 1993; Hofsten, 1984), crawling (Goldfield, 1989; Adolph, Vereijken and Denny, 1998), walking (Adolph, 1997; Gesell, 19439), communicate by language and imitate facial expressions (Fogel and Thelen, 1987; Meltzof and Moore, 1997) as well as understanding the physical properties of the world (Baillargeon, Spelke and Wasserman, 1985; Spelke et al., 1992; Thelen et al., 2001) and the outcomes that can be achieved by their own actions (Angulo-Kinzler; 2001). Later with development and

growth, humans learn culturally imposed motor behaviors, such as riding a bicycle, swimming, playing sports and musical instruments, writing, driving a car, etc. The mental and behavioral aspects of how humans learn new motor and cognitive behaviors, from infancy to late adulthood, have been extensively investigated for more than 100 years (James, 1914; Thorndike, 1911, 1921; McGraw, 1932; Adams, 1987; Thelen and Smith, 1994).

One old and recurrent topic that still captures much of the attention of researchers investigating human behavior is the phenomenon of stage transitions in motor learning. Famous psychologist William James (1890; 1914) was one of the first to suggest that the learning process takes place as a series of stage transitions in mental and behavioral performance. James emphasized the role of conscious and unconscious processes for behavioral control. In his view, when humans face a novel learning situation a conscious system takes behavior control through attentional mechanisms. On the other hand, an unconscious automatic system assumes control after continuous and repeated practice of the same task and behavior.

This very same problem of stage transition in motor learning was later pursued by contemporary researchers investigating different human behaviors and under multiple theoretical umbrellas (McGeoch, 1931; Fitts, 1964; Fitts and Posner, 1967; Adams, 1971; Schimidt, 1975; Adams and Bray, 1970; Gibson, 1969; Gibson, 1979; Newell, 1985; Ackerman, 1988; Gentile, 1972). Table 1 summarizes the features of the theories which are relevant for the purview of this thesis.

AUTHOR	THEORY	EXPERIENCE STAGES IN ACTION SELECTION		
Fitts (1964) Fitts and Posner (1968)	Transition of cognitive processes in motor learning	**Verbal-Cognitive** Use attention to establish the proper cognitive task set, sensory input and candidate actions	**Associative** Anticipation of sensory inputs and associated action and outcomes	**Autonomous** Performance becomes automatic, attentionless, fast and error-free
Ackerman (1989)	Transition of cognitive processes in perceptual learning	**Cognitive** Attentional demands for reasoning, spatial orientation, understand rules and strategies	**Associative** Acquisition of perceptual speed for quick response generation using vision or memory	**Autonomous** Association of perceptual-motor abilities
Newell (1985)	Learning to control and coordinate the multiple body segments	**Novice** Multiple joint-movements and co-contraction of muscles	**Advanced** Increased adaptability and reduced muscle co-contraction	**Specialized** Movements are energetically efficient and well coordinated in space-time
Hikosaka et al .(1999)	Coordinate representation in sequential procedural learning	**S-R formation** Formation of sensory-motor mappings based on single action single sensory input	**Spatial Coordinates** Formation of bonds linking actions using visuo-spatial coordinates	**Motor Coordinates** Use of a motor-based system after repeated practice of the same sequence
Savelsbergh et al. (2001)	Hierarchical acquisition of complex behaviors from elementary action blocks	**Simple Actions** Association of an action to a single sensory input and context	**Multiple Actions** Association of an action to multiple sensory inputs and binding	**Generalization** Flexible adaptation of learned actions to multiple contexts
Gentile (1972)	Manipulation of task difficulty to facilitate the learning-teaching process	**Simple Task** Acquisition of the representation of movement patterns in simple environments		**Complex Task** Adaptation of learned movement patterns in multiple and complex environments
Adams (1969)	Learning as error reduction by problem-solving and mental model formation	**Verbal-Motor** Use of over/covert speech to understand and acquire a representation of the task		**Motor** A nearly perfect representation of the movement is achieved and performance gets fast
James (1914)	Behavior control based on conscious and unconscious processes	**Conscious** Use of attentional process to guide learning and understanding of the task		**Unconscious** Cognitive/motor behavior becomes habitual

Table 1. Theoretical models of stages in action learning. The sequence of stages moves from left to right. Some theories propose the existence of at least three stages whereas others propose two main processes based on the initial and advanced characteristics of behavior performance. Other theories which are important for the work developed in this thesis will be touched upon in later sections.

2.1.1. The Fitts-Posner Three-stage Theory of Action Learning

At this moment, two of these classical theories deserve attention, the cognitive theory proposed by Fitts (1964) and Fitts and Posner (1967) and the error reduction theory based on reinforcement feedback advocated by Adams (1969). In the Fitts (1964) and Fitts and Posner (1967) theory, motor learning occurs in three stages and the cognitive processes used throughout learning are based mainly on the attentional demands and the subject experience with the task. In an early initial phase, humans first try to understand the task and the goals which are to be achieved. Therefore, in this phase it is important to establish a proper cognitive set which will be used during the task and that require high amounts of attention. The cognitive task set might be composed of the task rules, the task goals, the predominant sensory information and the actions that can be performed to achieve the goals. Since humans are still learning what can be done and how to solve the task, the performance in this initial phase is marked by a large number of errors and slow action selection and movement execution. The duration of this phase is dependent on the task complexity.

The learning processes moves on to the intermediate phase after the formation and acquisition of the cognitive task set. In the intermediate phase, humans become able to use the cognitive task set to make stronger stimulus-response (SR) associations by anticipating what action should be executed given that specific sensory information will appear. Attention still plays an important role in this phase, but the number of errors in performance decreases as humans become able to select proper actions with higher frequency. In the late phase of action learning human performance becomes automatic

with fast execution after repeated practice of the same actions. Different from the initial and intermediate phase when actions were selected based on external information, in the late phase actions are selected from internal memory representations, becoming less dependent on visual and attentional processes (Figure 1A).

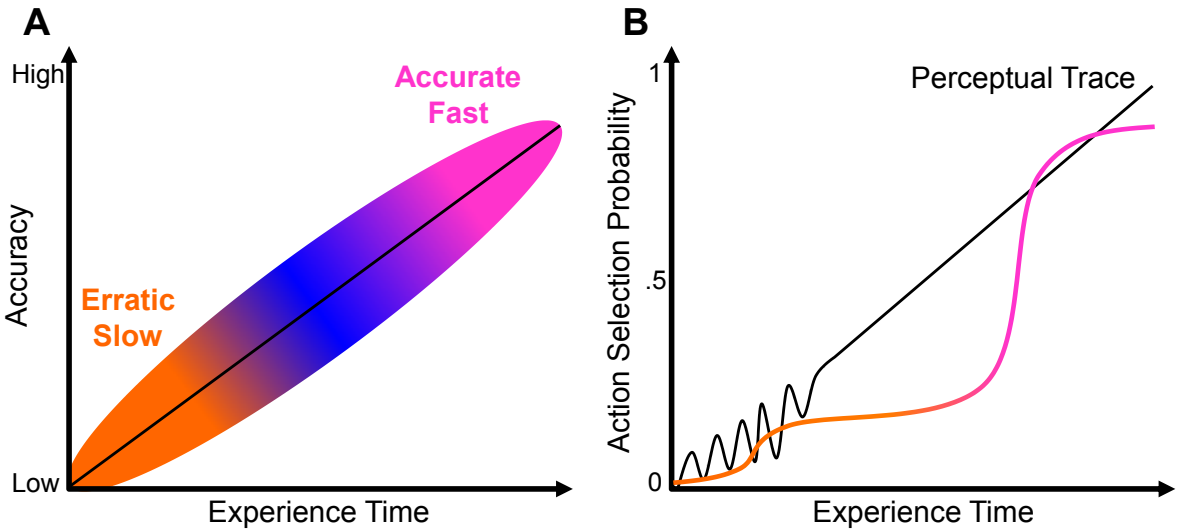

Figure 1. Schematic view of stage transition in motor learning and action selection probability distribution as a function of experience.
(A) Hypothetical depiction of performance accuracy improvement with learning based on the theories summarized in Table 1. The color gradient represents the multiple learning stages and transition across stages: initial (orange), intermediate (blue) and late (magenta). The black line assumes a linear but finite improvement in performance.
(B) Representation of Adams's (1971) theory on weak and strong action representations . The orange and magenta lines represent the memory trace for the same action with different selection probabilities through experience. The black line represents the perceptual trace with an initial low and noisy representation and high and accurate representation of the motor behavior after robust learning.

Although Fitts and Posner proposed the existence of an intermediate stage as humans learn to select actions, their idea is that the initial and intermediate stages share the same

cognitive processes. Consequently, their theory is in line with that of James (1890; 1914) and their proposal is that an attentional system is involved in controlling behavior in the initial and intermediate stages of learning, whereas an automatic system controls the performance of extensively practiced actions.

2.1.2. Adams's Reinforcement Learning Theory

In a 1971 Journal of Motor Behavior paper entitled "A closed-loop theory of motor learning", Adams proposed a learning theory to explain how humans might use reinforcements or punishments to learn motor actions. The theory came up as an opposing force to the influential works by Karl Lashley on internally guided movements (1951, 1917, 1948). One differential aspect between Lashley's and Adams's works is the type of motor behavior to be executed. Although Lashley's works consider learning an important process, the theory itself can be applied to the internal generation of any motor program initially triggered by a feedforward mechanism and which continues to be run by internal control processes. The theory also applies for the execution of sequential behaviors and decision making processes. On the other hand, Adams's theory tries to explain how graded slow movements will acquire a stable and stereotypical pattern as a function of practice, and how the movement pattern can be transformed by performance feedback.

Using engineering and mathematical concepts not explicitly described in the paper, Adam's proposed a 2-state 2-stage learning theory based on probability distribution for

action selection, formation of beliefs and reduction of uncertainty (Figure 1B). The 2-stage learning side of the theory proposes that in the initial stage of learning performance is controlled by a hybrid verbal-motor system where a verbal and motor systems work concurrently, but the verbal system takes predominant control. Later with experience, the late stage, the motor system assumes control and the verbal system is important mainly in cases where the task requires some sort of updating of the learned actions.

Adams (1971) considered action learning as a problem-solving process and argued that in such situations humans use either overt or covert speech in order to understand the task structure. In his view, the use of overt speech is part of an attentional mechanism. Therefore, speech and attention are essential factors in the early stage of learning. The late stage of learning in Adams concept is not different from previous proposals. Motor behavior becomes automatic or habitual after repeated practice and a memory of the movement pattern is stored somewhere in the nervous system. With experience, humans tend to make less use of verbal processes and actions can be executed with higher spatio-temporal precision, accuracy and speed.

A significant advance to the study of motor learning was Adams's proposal of an error correction mechanism as the 2-state part of his theory. Previous theories (James, 1890; McGeoch, 1931; Fitts and Posner, 1968) made important contributions by identifying changes in motor performance which were likely to be accompanied by changes in mental cognitive processes and set the path for contemporary research on motor learning with the use of advanced brain imaging methods, such as functional Magnetic Resonance Imaging (fMRI). However, these very same theories paid scant attention to the

mechanisms guiding learning and failed to explain how performance improves from trial to trial.

In Adam's 2-state learning theory, the error correction mechanism is composed by a *perceptual trace* and a *memory trace* (Figure 2). The memory trace is responsible for processes related to action generation such as retrieval, selection and execution. The perceptual trace is loaded after the memory trace chooses an action and is responsible to provide a measure of the goodness of that action by recognizing the quality of the movement performance by its error deviance from what was initially expected. Therefore, the error correction signal for the next movement results from the mismatch, that is, the difference between the actual motor performance and expected performance computed by the perceptual trace.

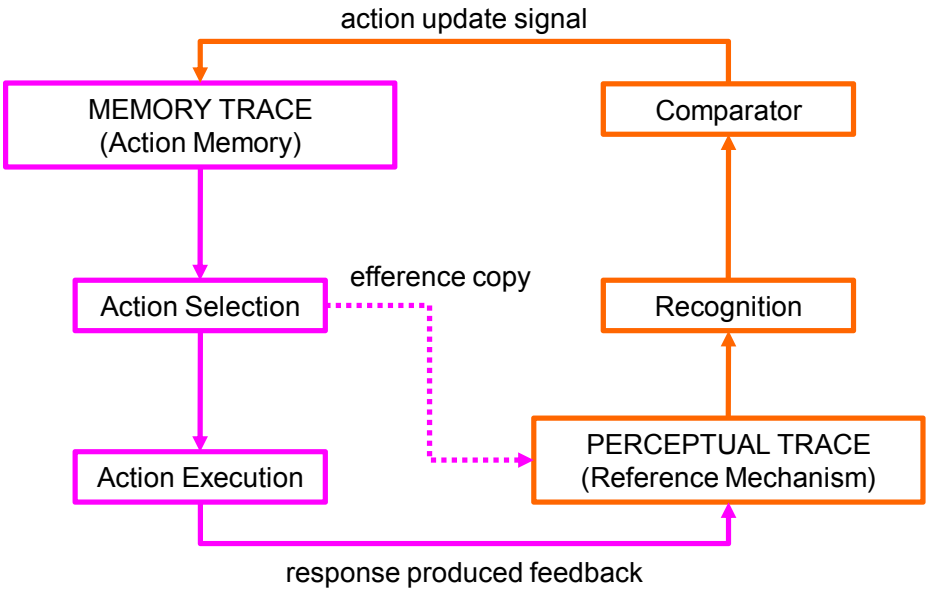

Figure 2. Schematic visualization of the 2-state theory of error reduction by reinforcement as proposed by Adams (1971). The memory trace is responsible for the storage of the motor repertoire, the selection and execution of actions. Once an action has been selected, an efference copy of the expected movement is loaded in the perceptual trace and this information is used as a reference to which the response produced feedback

should be compared to. If a mismatch between the expected and actual movement is found, the perceptual trace updates its representation and sends an error update signal for correction of the movement in the next future event.

Adams (1968, 1971, 1977) and his coworkers (Adams and Bray, 1970; Adams, Marshall and Bray, 1971) emphasized the use of this error correction system by studying how human subjects corrected error in well-learned motor behaviors, such as paired-associated verbal and arm positioning tasks. This situation raised strong criticism (Schmidt, 1975), since no details were provided on how the perceptual and memory traces would work for the learning of novel actions. However, Adams (1971) assumed that the perceptual and memory traces could be built up from scratch in the case of new learning. In the case of the memory trace, action selection probabilities are very sparse in the initial stages of learning, but the selection probability of an appropriate action and movement pattern increases with experience and the memory representation gets stronger. The same applies for the perceptual trace which is initially noisy and of little help for error reduction but it eventually learns a model for error correction from experience by trial-and-error (Figure 1B).

2.2. Neural Mechanism of Motor Learning

The extensive investigation of the motor learning processes and analysis of behavioral performance by the pioneers of the field paved the way to understand the brain

mechanisms that are involved in the different learning stages. Technological advances with the development of fMRI and positron emission tomography (PET) allowed researchers to scan the whole brain and record the neural activity as humans learn motor tasks (Jenkins et al., 1994; Jueptner et al., 1997a, 1997b; Toni et al., 1998; Grafton, Hazeltine and Ivry, 2002; Floyer-Lea and Matthews, 2004; Leherecy et al., 2005; Doyon et al., 1996; Doyon et al, 2002; Sakai et al., 1998; Jubault and Koechlin, 2007).

In addition, neuroanatomical and electrophysiological studies with animals, including monkeys and rodents, have also contributed to understand the cortico-cortical and cortico-subcortical roles in behavior (Saito et al., 2005; Miyachi et al., 1997; Lu, Hikosaka and Miyachi, 1998; Ljungberg, Apicella and Schultz, 1992; Barnes et al., 2005). The purview of this section is to provide an overview of the neural mechanisms of human motor learning and the similar processes that might be conserved in monkeys and rodents.

2.2.1. Neural Mechanisms of Motor Learning in Humans

The use of fMRI and PET has played an important role in the understanding of the brain mechanisms that are predominantly activated as humans learn new motor behaviors. In this section, the focus will be the problem of sequential procedural learning. Lashley (1951) was the first research to address the problem of sequencing motor and cognitive behaviors. Sequential motor behaviors include playing the piano, type writing, or any sequence of discrete or continuous movement patterns, whereas cognitive behavior

involve thinking, reasoning, langue, speech, etc.

Passingham and his colleagues (Jenkins et al., 1994; Jueptner et al., 1997a, 1997b; Toni et al., 1998) are among the first researchers to investigate sequential motor learning by humans. As many classical motor learning theories agree on the existence of initial and late stages in learning their principal contribution was to establish and differentiate the brain mechanisms that predominantly are activated in these stages as humans learned sequential finger movements.

In a series of experiments using PET, they tested humans subjects performing action sequences under several task conditions: new learning of sequences, attention to the next move and execution of well-learned sequences (Jueptner et al., 1997a, 1997b; Toni et al, 2002; Jenkins, 1994; Toni et al., 1998). The results of these experiments showed that distinct neural networks were predominantly activated during different stages of motor learning. Passingham's group found that certain brain areas are highly active in the initial learning of new motor sequences, including the dorsolateral prefrontal cortex (DLPFC), the anterior cingulate cortex (ACC), anterior part of the basal ganglia (BG), especially the head of the caudate nucleus and anterior part of the putamen. On the other hand, the analysis of the brain activity associated with the performance of well-learned automatic finger sequences showed that mainly the posterior supplementary motor area (SMA), posterior premotor cortex, posterior parietal cortex (PC), and posterior putamen were active (Figure 3).

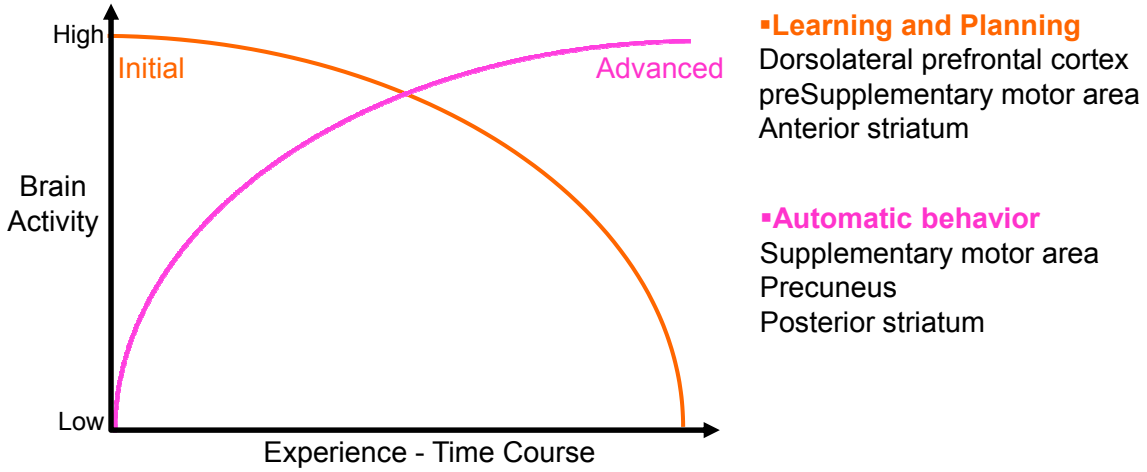

Figure 3. Schematic representation of the differential activity of multiple brain areas throughout learning. The orange line represents the activity of brain areas (on the right) that are highly active in the initial stage of learning but whose activity decrease with long-term experience; the magenta line represents the activity of brain areas which show an initial low activity profile which then gets stronger as the performance becomes automatic. Scheme based on Passingham's works.

An interesting finding in these experiments is that the areas that are highly active in the initial stage of learning decrease their activities as learning progresses, while brain areas that show low level of activity in the initial period of learning increase their activities after repeated practice and achievement of automatic behavior. These results also suggest that human behavior might be under the control of distinct neural networks, a cognitive network for the learning and planning of motor behaviors, and a motor network for implementing well-learned motor actions.

2.2.2. Neural Mechanisms of Motor Learning in Monkeys

Anatomical lesions and single neuron electrophysiological recording studies have also been successfully applied to establish with higher validity the biological mechanisms of motor learning. In another famous series of experiments on motor sequence learning with monkeys and humans, Hikosaka and his colleagues (Rand et al., 1998; Rand et al., 2000; Sakai, Kitaguchi and Hikosaka, 2003; Sakai, Hikosaka and Nakamura, 2004; Hikosaka, et al., 1995; Hikosaka et al., 1999; Hikosaka et al., 2002; Nakamura, Sakai and Hikosaka, 1998; Miyachi et al., 1997) have found that causing reversible lesions by injection of muscimol into distinct brain areas impairs the learning of new motor sequences and retrieval of well-learned ones.

Hikosaka's group investigated how lesions in the SMA and BG affected monkeys behavior while performing a sequential finger movement task. Muscimol injections were applied in separate experiments in the anterior and posterior parts of the BG and SMA. This separation was based on anatomical connections that the BG make with prefrontal areas including the SMA and DLPFC. The posterior BG receives its main cortical afferents from the posterior region of the SMA, whereas the main afferent inputs on the anterior BG arrive from the anterior region of the SMA, called preSMA, and from the DLPFC.

When muscimol injections were applied to the anterior BG or the preSMA, monkeys had a great impairment in the learning of new finger motor sequences. On the other hand, lesions made in the posterior BG or in the SMA caused detrimental effects in the

performance of motor sequences that had been extensively learned before lesions. Electrophysiological recordings of single neurons supported the results of these lesion experiments. Neurons the preSMA and anterior BG showed preferred activity when monkeys performed new movement sequences, while neurons in the SMA and posterior BG showed higher activity if monkeys executed old motor sequences (Figure 4).

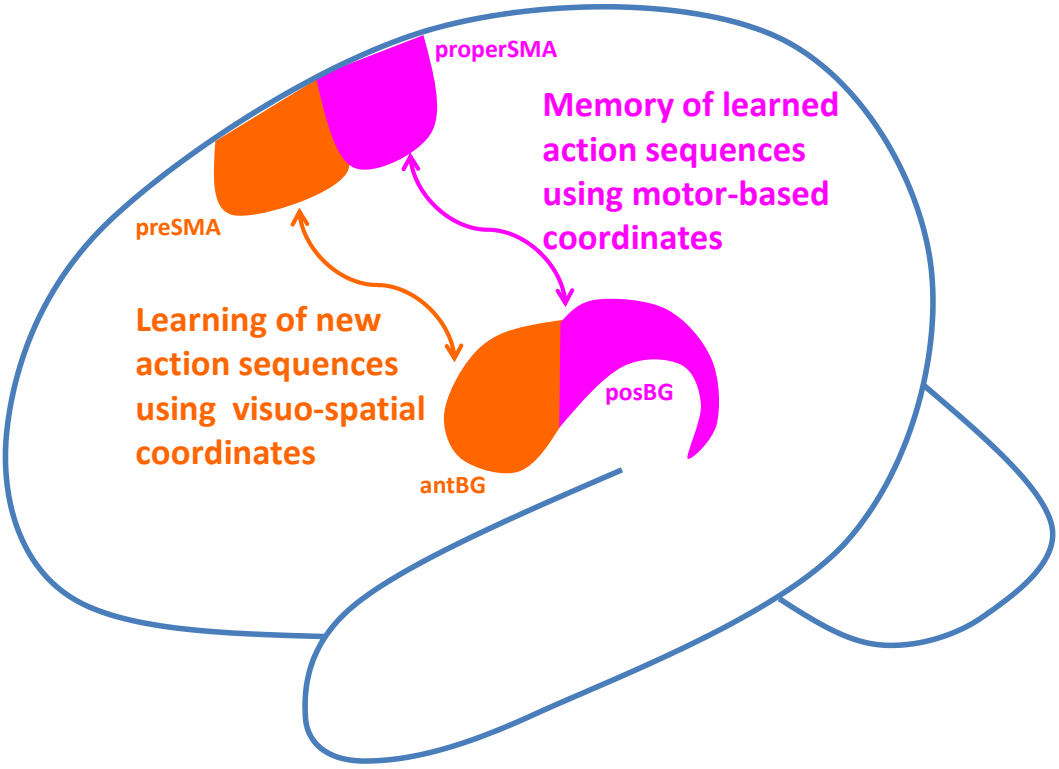

Figure 4. Schematic representation of the networks linking part of the prefrontal cortex, the supplementary motor area (SMA), with the BG. The anterior part of the SMA, the preSMA is connected with the anterior part of the BG, antBG, and together are important in the learning of new motor sequences. The posterior SMA, properSMA, sends fibers to the posterior BG, posBG, and participate in the implementation and execution of well-learned motor sequence memories. Adapted from Hikosaka et al., 1999.

2.3. Parallel Anatomical Neural Networks for Behavior Control

The results of the imaging studies with humans and those of the lesion and single neuron electrophysiological recordings with monkeys suggesting that multiple neural networks control behavior in different learning situations are strongly supported by neuroanatomical evidence showing parallel segregated closed-loops linking cortical and subcortical brain areas with recirprocal connections (Alexander et al., 1986; Alexander and Chutcher, 1990; Bostan, Dum and Strick, 2010; Kelly and Strick, 2003; Haber, 2003; Selemon and Goldman-Rakik, 1985; Eblen and Graybiel, 1995; Flaherty and Graybiel, 1993; Middleton and Strick, 2000; Glickstein and Doron, 2008; Dum, Li and Strick, 2002; Ramnani, 2006; Akkal, Dum and Strick, 2007; Nakano et al., 2000)

In a seminal paper Alexander, DeLong and Strick (1986) proposed a theory on the existence of parallel segregated loops linking cortical and subcortical areas. In that paper, emphasis was given to the bidirectional fiber connections linking the prefrontal cortex (PFC) with the basal ganglia (BG). In subsequent studies, the use of a novel technique based on anterograde and retrograde transneuronal virus herpes transportation, the same group demonstrated that parallel pathways connections linking the PFC to the cerebellum (CB, Dum and Strick, 2003) and more recently evidenced has been found showing anatomical connections between the CB and BG (Hoshi et al., 2005). These multiple pathways are likely to be involved in the processing of distinct types of information, such

as motivational, emotional, cognitive and executive functions as well as motor control, including somatotopic representations.

Alexander et al., (1986) suggested the existence of five main parallel segregated PFC-BG circuits. The details of these circuits are illustrated in Figure 5. The connections linking the PFC with the BG are arranged as a gradient distributed in anterior-posterior and medial-lateral directions and reach the striatum, the input regions of the BG, the caudate nucleus and putamen (Alexander et al., 1986; Haber, 2003; Frankle, Laruelle and Haber, 2006; Selemon and Goldman-Rakic, 1985; Selemon and Goldman-Rakic, 1988).

The ventral part of the PFC, the medial orbitofrontal cortex (MOFC) and anterior cingulate cortex (ACC), projects to the ventral striatum (nucleus accumbens, NAc) and to the anterior and medial areas of the head of the caudate nucleus and anterior putamen. The dorsolateral prefrontal cortex (DLPFC) projects mostly to the anterior BG and middle of the caudate nucleus. The motor cortex, including the SMA, preSMA, primary motor cortex (M1), somatosensory cortex and posterior parietal cortex (PPC) project mainly to the posterior part do the putamen and the tail of the caudate. The newly found organization of the neural anatomical connections demonstrated by these studies implicate the participation of the BG in non-motor cognitive behaviors and changed the previous dominant view of its involvement in motor control.

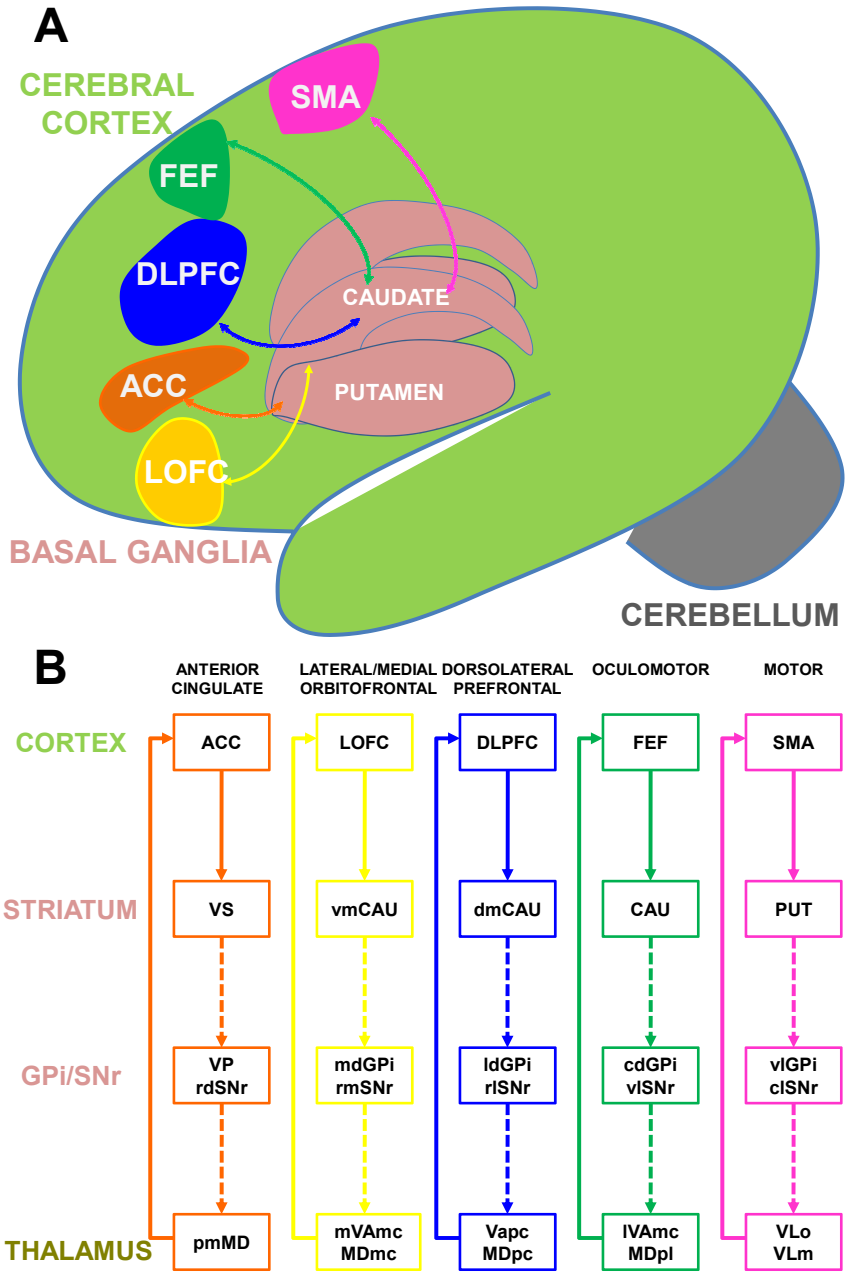

Figure 5. Schematic representation of the five PFC-BG circuits as initially proposed by Alexander et al., (1986).
(A). Distinct regions of PFC that have bidirectional connections with the BG.
(B). Detailed view of the five parallel PFC-BG closed-loops. The top layer represents the structures in the cerebral cortex; the second layer shows the input structures of the striatum; the third layer represents the output structures of the striatum and the fourth layer represents the connections back to the cerebral cortex through the thalamus. DLPFC, dorsolateral prefrontal cortex; ACC, anterior cingulate cortex; SMA, supplementary motor area; FEF, frontal eye field; LOFC, lateral orbitofrontal cortex; VS, ventral striatum; vmCAU, ventromedial caudate; dmCAU, dorsomedial caudate; PUT,

putamen; VP, ventral pallidum; GPi, globus pallidus internal segment; SNr, substantia nigra pars reticulata; MD, mediodorsal nucleus; VA, ventro anterior nucleus; VL, ventrolateral nucleus.

Another important feature of the BG system is its prominent connections with the dopaminergic system which is composed of the ventral tegmental areas (VTA) and substantia nigra pars compacta (SNc; Haber, 2003). Dopamine (DA) is a neuromodulator known to have important roles in instrumental learning by associating a motor behavior with an expected reward that follows an action execution and in other learning and memory processes. It also plays roles in other motivational aspects such as guiding attention to salient events or novel stimulus information in the environment (Dommett et al, 2005; Redgrave and Gurney, 2006). It has been suggested that higher concentration of dopamine in the BG and cerebral cortex strengths emotional processes such as making bonds between a mother and child (Bartels and Zeki, 2004) and romantic love (Fisher, Aron and Brown, 2006; Bartels and Zeki, 2000).

Through a series of anatomical studies, Haber and colleagues (Fudge and Haber, 2000; Haber, Fudge and McFarland, 2000; Haber et al., 2006; Lynd-Balta and Haber, 1994a, 1994b, 1994c) found out that distinct regions of the dopaminergic systems make bidirectional spiral-like connections with distinct areas of the BG. The VTA sends to and receives connections mostly from the ventral striatum including the nucleus accumbens (NAc) and anterior and ventral parts of the caudate nucleus (head of the caudate) and putamen. The SNc can be subdived into two tiers, namely: the densocellular and the cell columns. The SNc densocellular ventral tier projects to and receives inputs from the dorsomedial striatum and the cell columns tier sends fiber to and receives from the

posterior striatum (Figure 6). An interesting aspect of these dopaminergic connections to the striatum is that the anterior BG also projects back not only to the VTA but also to the SNc densocellular tier, and the dorsomedial BG projects back not only to the SNc densocellular tier but also to the cell columns tier. Haber (2003) suggested that this spiral-like connection is a mechanism that the motivational system (medial orbitofrontal cortex and VTA) can exert influence on the cognitive system (DLPFC and dorsomedial BG). The same holds for the cognitive system which can modulate the functions of the motor pathway (SMA and posterior BG).

Strick and his group (Akkal, Dum and Strick, 2007; Dum, Li and Strick, 2002; Dum and Strick, 2003; Kelly and Strick, 2003; for an insightful review on the PFC-CB connections and functions see Ramnani, 2006) also found distinct bidirectional fiber connections linking areas of the PFC to the cerebellum (CB). The DLPFC sends fiber connections to and receive from the lateral hemispheres of the CB, including areas Crus 1 and Crus 2. On the other hand, the primary motor cortex sends to and receives fiber connections from the anterior inferior posterior regions of the CB. These results demonstrate that the CB and BG are in a position to influence both motor and higher order cognitive processes. The precise functions that these different modules, PFC, CB and BG, play together in motor and cognitive information processing, such as learning and memory, are still unclear. However, that are many indications that the BG might be important for instrumental and reinforcement learning (Graybiel et al., 1994; Horvitz, 2009; Doya, 2007), the CB in storage of motor memories and internal models of the environment and body (Kawato, 1999; Wolpert and Kawato, 1998) and the PFC in

working memory, reasoning, mental simulation and prediction of future events (Doya, 1999; Samejima and Doya, 2007; Glascher et al., 2010).

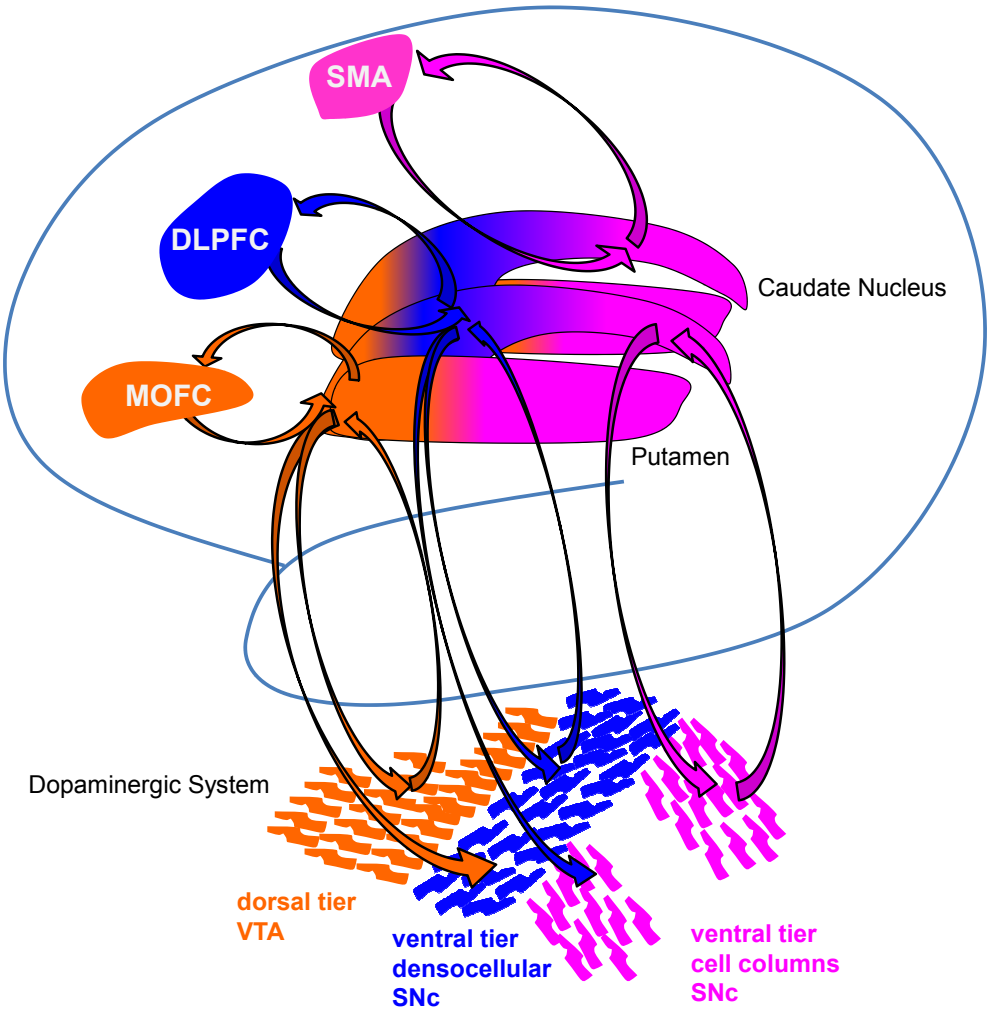

Figure 6. Topographical connections linking the PFC-BG-DA system. The BG (center of the brain) is represented as a color gradient representing the predominant input connections they receive from and send back to the PFC. The spiral connections linking the DA system with the BG provide a way that information between multiple segregated closed-loop connections between the PFC and BG can be integrated. VTA, ventral tegmental area; SNc, substantia nigra pars compacta; SMA, supplementary motor area; DLPFC, dorsolateral prefrontal cortex; MOFC, medial orbitofrontal cortex.

2.4. Contemporary Theories of Action Selection and Learning

The pioneers in the study of motor behavior have been highly influential by providing the results of elegant experiments showing the fine details of changes in motor performance as humans learn novel motor actions. They also synthesized these results in an attempt to propose theories capable to explain the several cognitive and motor processes involved in learning in general. These prevailing ideas, including the stage-wise thinking of motor learning, set the path for contemporary researchers who could use new technologies to investigate the changes in brain processes as learning takes place.

Recently, contemporary researchers started to use computational models in order to understand in a more accurate manner what information is important for the generation of motor commands, the computational architecture responsible for information processing, the methods used for action selection and motor control as well as the associated brain mechanisms. These computational approaches have also been used to explain how the normal neural and psychological mechanisms of cognition and motor control are impaired in neurological and psychiatric conditions (Blakemore, Wolpert and Frith, 2002; Frith, Blakemore and Wolpert, 2000).

This section will provide a brief overview of the prominent computational theories of motor control and learning, including Reinforcement Learning (RL), how these theories have been helpful in the understanding the constituent components of normal and

disturbed control processes as well as their related counterparts in the brain.

The use of computational approaches to the study of motor behavior has been helpful in understanding the main components and information processing requirements for the generation of motor outputs and control of movement execution. Wolpert and Flanagan (2010) have proposed that these components can be classified into four groups: efficient gathering of sensory information, making a model of the environment and task, establishment of a control strategy, higher-level skill acquisition. An outline of these components will be described below.

Efficient gathering of information is related to learning how to extract task-relevant sensory information. This information extraction is dependent on the sensory modality the information access the central nervous system, such as visual, haptic, auditory, etc. The generation of a motor response might also be influenced by the sequence of sensory events. In a predictable and well-learned environment, humans become able to anticipate a sensory event and prepare a motor response even when the sensory information is still unavailable. On the other hand, in complex task with unpredictable and uncertainty features, performance might be subject to the minimization of the uncertainty as humans acquire search strategies in order to focus their attention on relevant and important sensory information.

Learning a model of the task, on the other hand, refers to the different structural properties that change from one task to the other. For instance, playing soccer requires running, dribbling, passing and kicking a ball, whereas swimming requires learning of

head side movements for breathing, constant synchronization of legs and arms and turning by submerging and a later emerging on the water surface. On the other hand, other tasks share similar structures and very often subtle adaptations should be necessary. For instance, a larger racket is need for playing tennis whereas smaller racket is needed for playing badminton. However, these two sport modalities differ on the size of the court and height of the net in the middle of the court. Therefore, spatial adjustments and force control mechanism have to be modulated for dislocation within the large and small spaces, and for controlling the racket, respectively.

A predictive strategy for motor control is one among many other methods. The prediction model can be used for the generation of motor commands in anticipation of sensory events. A class of predictive models, called forward model, can be used by mental simulation for the prediction of future states reached by hypothetical actions. These models can generate fictive states and actions, evaluate the goodness of taking one action but not the other, predictive the future states if a given action is chosen for execution and the anticipation of possible and sudden sensory outcomes that might results from the execution of a motor behavior or other sources of changes in the environment.

In more recent works using computational motor control approaches, the three components described above have been expanded to explain other cognitive processes such as decision making and social interaction. It has been suggested that forward models might be used in social interactions, such as the prediction of the outcomes of the behaviors of others. These predictions might be influenced by the sensory information

coming from the body gestures, speeches or sequence of sensory inputs combined from a single agent of multiple agents interacting in an active or passive manner. Also, a recent theory of motor control has incorporated the feedback generated from motor actions as an important aspect for improvement and efficient generation of motor commands (Todorov, 2004). This theory, called optimal feedback control, defines the motor behaviors have a cost and the goal is to reduce this cost and saving of energy. In the initial stage of learning, the cost might be very high given the task novelty and absence of parameters for the optimal generation of movements. With experience, however, humans learn to reduce the errors in the movement execution. An important aspect of this theory is that, error correction should happen only for those aspects of the movement that are deleterious to the task.

Doya (1999, 2000) proposed that three main classes of learning algorithms might be implemented in the human brain depending on the types of feedback signals used for error correction: unsupervised learning, supervised learning and reinforcement learning. In unsupervised learning, no feedback is provided, but only a set of input sensory data and the goal of the system is to learn the statistical properties of the input and construct a map of the sensory input with possible output signals. In supervised learning an error vector is provided and corrections and updates are made to the input-output mapping. Reinforcement learning uses a scalar evaluative reward signal to map states to actions. The goal a reinforcement learning agent is to maximize the maximum long-term expected reward. Doya (1999, 2000) suggested that these learning algorithms might be implemented in distinct brain regions: unsupervised learning in the cerebral cortex, supervised learning in the cerebellum and reinforcement learning in the basal ganglia.

2.5. Unsolved Problems in Action Selection and Learning

It is well accepted that the learning process might occur as a transition from an initial erratic and slow performance stage to an advance fast and accurate performance stage. However, there has been some evidence suggesting that an intermediate planning/predictive stage might follow the early learning stage (Hikosaka et al., 1999; Flanagan et al., 2003; Savelsbergh et al., 2001; Fitts and Posner, 1969). These studies suggest that the initial learning stage might important for the acquisition of the structure and model of the task, including the predominant sensory input, the available actions and possible outcomes. Other somatosensory-motor processes are also part of the initial stage, such as learning to control a limb a sequence of fingers. In addition, learning itself might be driven by emotional/motivational factors that keep the agent engaged at solving the task.

As learning proceeds and enters the intermediate stage, the model acquired initially can now be used in a more efficient manner for planning, mental simulation, and prediction of future sensory states, motor outcomes, and evaluation of the goodness performing an action of multistep actions. If the task requires the acquisition of stereotypical movements with fixed kinetic, kinematic and speed profile patterns, then learning proceeds to an advanced learning stage. It is not clear, however, whether this sequence of learning stages occurs as a serial process in which the next comes to take cover behavior control only after the preceding learning process or whether the multiple learning processes operate in parallel.

Another problem in motor learning is related to the phenomenon of transfer of learning (TL). Theories of TL suggest that knowledge acquired in past learning situations can be used to guide and speed up learning in other decision making or other learning situations. TL of learning is classically subdivided in two main types: task-based or knowledge-based. Task-based TL is related to the execution of the same motor response under different situations, such as reaching for a glass of beer or reaching for a cup of coffee. Knowledge-based TL refers to the use of previously acquired knowledge for the generation of novel motor responses. A problem with TL theories is the assumption that once a motor or knowledge representation is acquired and stored in the brain, the mechanisms initially used for their learning are no more required. This contradicts recent studies showing that the use of knowledge representation, such as chess rules, re-engages areas associated with learning and planning, even when the task does not require learning or formation of stereotypical motor responses.

The early human brain imaging studies focused mostly at identifying the brain regions that were predominantly activated in distinct learning stages. These studies, however, failed to identify what functions or computational operations the brain regions play for learning, planning or generation of reactive habitual behaviors. Recently, with the introduction and design of experiments following RL learning principles, is has been possible to map basic computational processes to certain brain regions. For instance, it is now know that the dopaminergic neurons the SNc might compute the difference between the expected reward outcome and predicted reward outcome, a computational process called TD-learning. However, there have been suggestions that more elaborated action selection strategies, such as Model-Free (MF) and Model-Based (MB) might be utilized

by humans and have a neural correlate in the brain. In addition, recent work (Shadmehr and Krakauer, 2008) has proposed a computational architecture that can generate motor behaviors by implementing complex action selection strategies or simple reactive behaviors. There has been, however, no direct test of these hypothesis and it is not clear whether such computational processes are actually carried out in the human brain.

Lesions to a specific brain area can cause damage to a specific cognitive process. For instance, if the hippocampus is lesioned, humans become unable to make new memories and if the posterior region of the BG is removed, a deficit in the retrieval of learned motor behaviors is observed. On the other hand, lesions to other brain areas can also cause similar impairments in making or retrieval of memories. Therefore, it is not known whether the brain mechanisms involved in action selection are implemented locally by a single brain area or in neural networks composed of multiple interconnected of segregated brain areas.

Another intriguing and still unresolved issue is the amount, type or set of information that a certain brain region needs to carry out efficient operation. Also, it is not know what the actual computational process a brain region uses for information processing. Since many brain regions interact in parallel closed-loops but also interact with each other through feedback integrative loops, the direction of information flow from one region to the other, and which region and under which circumstances starts the processing of information is still to be elucidated.

CHAPTER 3

Evidence for Model-based Action Planning in a Sequential Finger Movement Task

3.1. Introduction

Humans and other animals have an amazing capability of flexibly and adaptively selecting actions according to dynamically changing task demands and environments. Novel motor behaviors can be learned from scratch by trial-and-error or by utilizing knowledge from past experiences that facilitate or speed up action acquisition.

In Reinforcement Learning (RL), a computational theory of adaptive optimal control (Sutton & Barto, 1998; Doya, 2007), two methods for action selection and learning have been proposed: model-free (MF) and model-based (MB) methods. In the model-free method, the evaluation of states and actions and according action policy are learned by trial-and-error from actual exploratory experiences. On the other hand, in the model-based method, an internal model that predicts the results of hypothetical actions is given

or learned, and then use the consequences of the simulated hypothetical actions for evaluating states and actions and for selecting an action. The aim of this study was to investigate whether and when human subjects use model-free and model-based methods in learning sequential motor behaviors and to elucidate how those learning methods are realized in the brain.

Classical theories of motor skill learning had put forward stage-wise transitions of motor behaviors and cognitive operations behind them (Fitts & Posner, 1967; Adams, 1971; Newell, 1985; Ackerman, 1988). Fitts and Posner (1967) proposed a serial three-stage model of motor learning based on simple discrete and rhythmic target reaching tasks. In this model, learning proceeds from an early cognitive/attentional stage where relationships between environmental cues and the control and regulation of movements are learned. In an intermediate associative stage, the learned cue-action relations are used to plan, refine and anticipate an action consequence. Finally in the late autonomous stage, movement performance becomes automatic with minimal attentional control. However, these classical theories cannot explain the mechanisms leading to stage transition, neither the cognitive operations nor the neural correlates operating as learning goes on.

A solution to this open debate has been put forward by Hikosaka and colleagues (Hikosaka et al., 1999; Hikosaka et al., 2002) in a series of elegant sequential finger movement studies using extensive psychophysical and electrophysiological recordings in monkeys and brain imaging techniques with humans. It was shown that distinct brain regions become preferentially activated in specific stages of learning. The anterior cortico-basal ganglia network linking the left dorsolateral prefrontal cortex (DLPFC), the

pre-supplementary motor area (pre-SMA) and the anterior basal ganglia (BG) was more active in the learning of new action sequences, whereas the posterior cortico-basal ganglia network including the proper-SMA, the primary motor cortex (M1), and the posterior BG and parietal cortex (PC) was activated when well-learned action sequences were performed.

These results were successfully replicated in a computational study (Nakahara, Doya & Hikosaka, 2001) which confirmed the previous suggestions that the anterior network learns faster by using visuo-spatial coordinates and that the posterior network learns slowly but forms robust memories using body-based motor coordinates. The authors concluded that different controllers, realized by distinct brain regions, are responsible for the organization of human behavior in different contexts.

Recent neuroanatomical studies found the existence of multiple parallel segregated loop channels through the prefrontal cortex, the parietal cortex, the basal ganglia and cerebellum (Alexander, DeLong and Strick 1986; Middleton and Strick, 1994, 2000; Haber, 2003) and that these loops process different types of information related to cognitive, emotional and skeleto-motor functions providing further support to the view that the brain contains multiple controllers.

Based on a large body of anatomical, physiological, and computational studies, Doya (1999, 2000) proposed that the cerebral cortex, the cerebellum and the basal ganglia can be conceptualized as structures specialized for implementing unsupervised, supervised and reinforcement learning algorithms, respectively, and hypothesized that a global

network combining the cortico-cerebellar loop and the cortico-basal ganglia loop can realize model-based reinforcement learning using internal models learned by supervised learning in the cerebellum.

Albeit increasing evidence on the existence of distributed controllers in the brain, there has been little effort and no direct test of whether and how humans utilize different control strategies for action selection, and where in the brain they are implemented. In this paper, we hypothesize on the existence of multiple control networks in the brain that learn concurrently using different learning strategies, and that these distinct networks take dominant role in motor control depending on the progress of learning, extent of experience and prior knowledge, leading to apparently stage-wise transitions in the learning of novel motor behaviors.

More specifically, we consider two model-free and one model-based reinforcement learning strategies as candidate algorithms in the learning of action selection. There are two main subtypes in model-free reinforcement learning algorithms: those based on the action-value function, such as *Q*-learning (Watkins and Dayan, 1992) and SARSA, and those based on direct action policy representation, such as actor-critic and policy gradient. In the action-value based methods, the action policy (or control law) is given implicitly from the action value function Q(s,a), which estimates the subsequent rewards by taking an action a at a state s. In the direct policy-based methods, the action policy P($a|s$) is memorized and gradually updated based on the difference of the predicted and actual rewards. The direct policy-based methods have the virtue of robust performance and simple real-time computation once an appropriate policy has been learned, albeit often

slower learning. The model-based method assumes that an internal model that predicts the resulting next state *s*' of a hypothetical action *a* at a sate *s* has been acquired and an action is selected by internal simulation of potentially multiple steps of actions. Assuming that the internal model is correct and the rewarding state is known, a model-based method allows learning from small number of actual actions, despite need for more time and working memory load for real-time operation.

In order to test behaviorally whether these different methods are used under what condition and to elucidate by functional brain imaging where in the brain they are implemented, we developed a new task paradigm called *grid-sailing task* and carried out an experiment with 16 human subjects. In this task, a subject performs sequential finger movements to move a cursor from a start position to a goal position on a 5x5 grid environment. The cursor can move to only three of eight neighbor grids so that reaching from the start to the goal often requires zig-zag movements, like in sail boat navigation. The subjects did not know the mapping between the three keys and three movement directions, which we call the *key mappping* (KM), so that they had to learn appropriate sequential finger movement from the scratch in order to maximize the score given according to the steps required to reach the goal.

We now elaborate on how concurrent learning by such model-free and model-based algorithms can result in stage-wise manifestation of different control behaviors (Figure 1). In the initial stage of learning, as the subject do not have an internal model for predicting the result of an action, so that only model-based algorithms cannot be used. Among model-free algorithms, action value-based learning is likely to take a dominant

role by increasing the value of state-action pairs that lead to successful goal reaching. In the intermediate stage, as the subjects acquire the internal model, the model-based algorithm should take a dominant role in guiding the actions, despite the need of more time for internal simulation. Finally, in the late learning stage in a fully predictable environment, the learned policy of model-free algorithm would play a major role as it requires less computation in real time, allowing quick and efficient execution of motor responses.

We assume a mechanism for selecting or combining the outputs of the multiple learning algorithms depending on their extent of learning, for example, by multiplicative integration of action selection probabilities. Note that how quickly the model-based algorithm can become usable is task-dependent; in tasks in which the result of an action is easily predictable, model-based algorithm may be used from an early stage while in some tasks model-based algorithm is not practical.

In order to reproduce the subject's behaviors and brain activity at different learning stages during a limited scanning time, each subject was pre-trained with particular combinations of key maps and start-goal positions and performed the task under three different conditions. In condition I, subjects used a new key map, in which we expect the use of the model-free, action value-based algorithm. In condition II, subjects used pre-learned key maps for new start-goal positions, in which we expect the use of a model-based algorithm, if enough time for internal simulation is provided. In condition III, subjects used pre-learned combinations of key maps and start-goal positions, in which we expect the use of learned policy of a model-free algorithm paradigm.

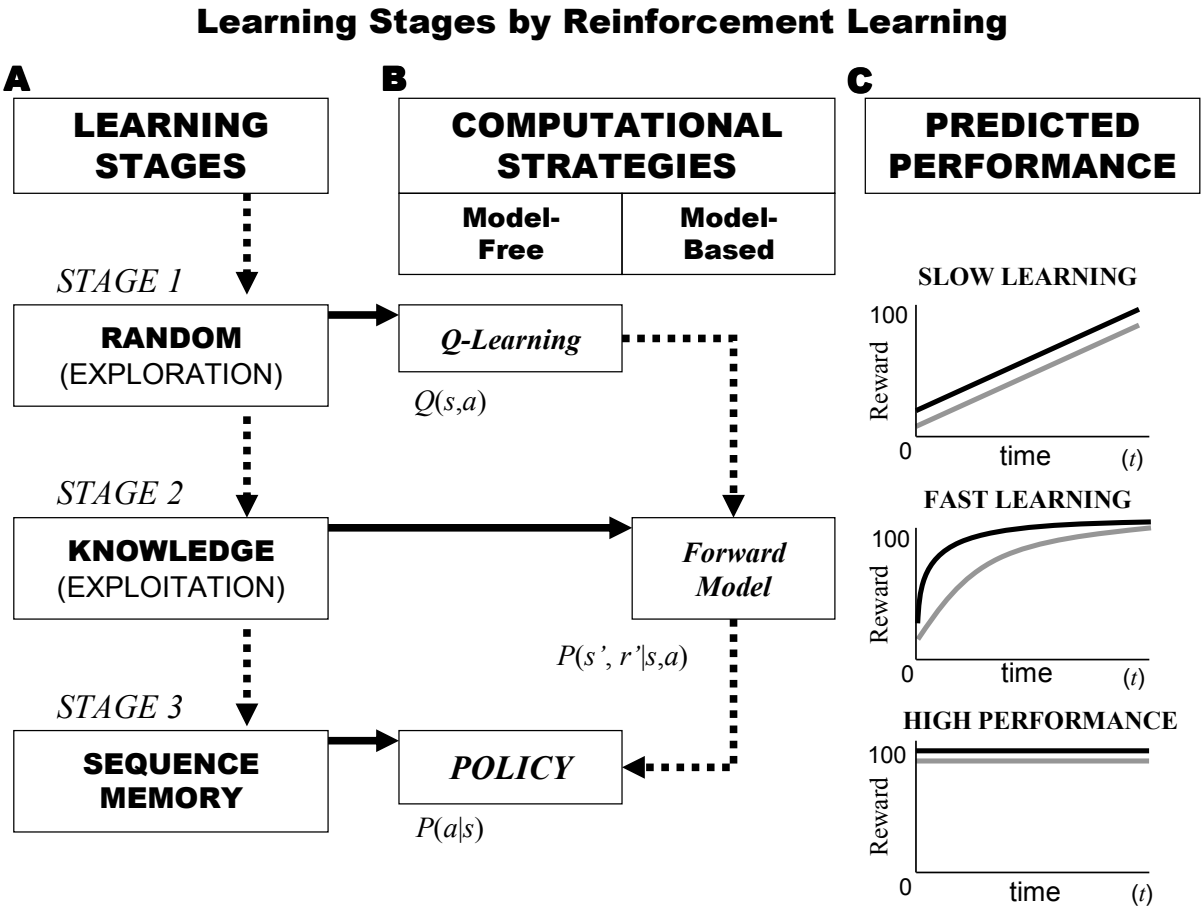

Figure 1. Schematic representation of a three-stage-three-system model for action selection and learning.
(A) Learning stages according to RL.
(B) RL algorithms for action selection and learning. The horizontal arrows represent the parallelism of the systems. The broken arrows the different represent the stage-wise transition in learning and the selection strategies predominantly employed in each stage.
(C) Expected performance under the control of each learning system. The black lines indicate the expected performance in trials with a delay period, and the gray lines indicate the expected performance (performance in delay start, black line, and immediate start conditions, gray line).

3.2 Methods

3.2.1 Participants

Sixteen healthy right-handed subjects (4 female; age 21-39, 26 ± 5 yrs), non-musicians, with normal or corrected to normal (by contact lenses) vision participated in this study. Subjects gave written informed consent and the experiment was approved by the ethics and safety committee of Advanced Telecommunications Research Institute International (ATR). Subjects received payment for their participation in this study.

3.2.2 Apparatus

In the training session, subjects comfortably sat at a table approximately 60 cm from the computer monitor. A regular desk keyboard was used for finger key-pressing. Subjects used the right index, middle, and ring fingers and pressed the ‘v’, ‘b’, and ‘n’ keyboard buttons, respectively (Figure 2A). Stimulus generation and behavioral data recording were done by custom-made programs written in MATLAB (version 7.0), using Psychophysics Toolbox functions (PTB-3). The test session was performed inside the fMRI scanner (3T Siemens, Magnetom Trio). A response optic button box was used obeying the same order of finger-key association (described above). Subjects received a

short practice to get accustomed to the button box before entering the fMRI scanner. In this paper we report only the behavioral results.

3.2.3 Task Procedures

Subjects were given verbal instruction of the task rules before the training session commenced, and received short verbal feedback orientation during the first training trials. The computer displayed the 5x5 grid (10 cm x 10 cm) with a red fixation cross (FC) on top of the grid center square matching the monitor center position. A trial started with the simultaneous presentation of a pair of start-goal (SG) positions, and a cursor (black triangle) which appeared on top of the start position. The color of the start position specified the start time condition: if the color was green, subjects were instructed to immediately start responding by performing key presses, while a red color indicated a delayed start condition when subjects had withhold key presses until go cue was signaled by the color change from red to green after a delay of 4 to 6 seconds. The color of the goal position was always blue. After the go signal, either for immediate of delay start trials, subjects had a maximum response time of 6 seconds, which was not explicitly taught to the subjects. During the response period, subjects had to execute sequential finger key presses to move the cursor from its initial start position to the goal position. Immediate visual feedback of cursor motion was provided after a button had been pressed. When the cursor was on the edge or corner of the grid, pressing of certain keys could result in an invalid move out of the grid. In such a case the cursor remained at the same position, and its color blinked to indicate that an impossible action had just been

performed. At the end of a trial, performance feedback information was displayed for 2 seconds and the next trial started after an inter-trial interval (ITI) of 3 to 5 seconds. Figure 2E shows a schematic representation of the sequence of trial events.

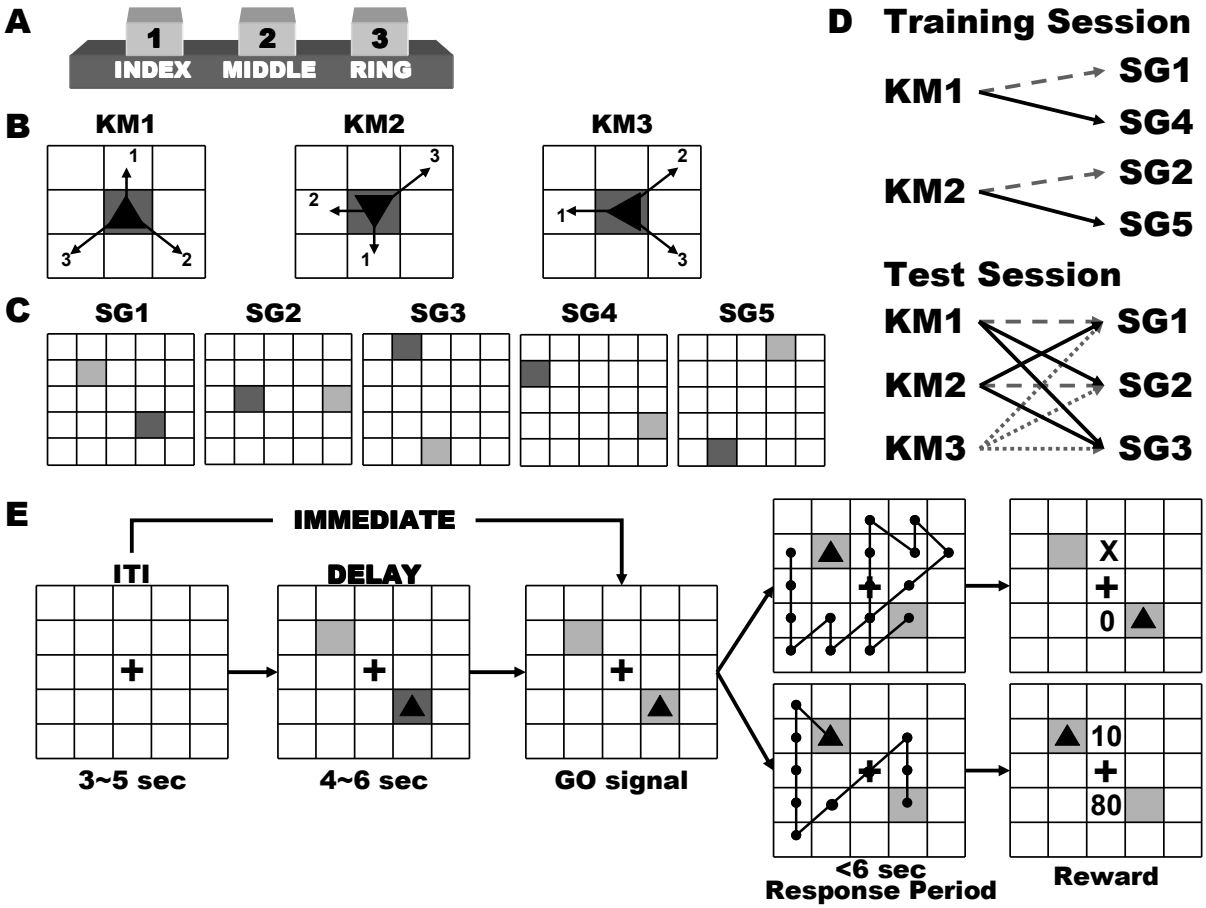

Figure 2. Task design and sequence of trial events.
(A) Button box and corresponding finger of the right hand.
(B) Key-Mappings (KM). Each KM is associated with a three-motion direction rule, and the numbers indicate the cursor motions and respective fingers.
(C) Start-Goal (SG) pairs. The start position is indicated by the dark gray color square, and the goal position by the light gray color square.
(D) KM-SG sets in the training and test sessions (Condition I: dotted gray lines; Condition II: black lines; Condition III: dashed gray lines). The KM-SG sets (dashed arrows) in the training session were used to provide subjects with more experience in finding all possible motion directions associated with each cursor KM.
(E) Representation of a sequence of trial events. ITI: intertrial interval.

A single trial had three possible endings according to a subject's performance in the response period: a) successful trial with performance of an optimal action sequences (OAS, shortest path to goal); b) successful trial with performance of non-optimal action sequences (NOAS); c) error trial when subjects could not reach a target goal within the maximum response period.

A point system was used as reward: 100 points for an OAS; a discount of -5 points for each excessive key-press in NOAS trials, and 0 (zero) point in error trials. Performance feedback was displayed on top of the grid at the end of a trial. In successful trials, a number appearing above the FC indicated the number of key-presses executed and the one below the FC the total reward acquired for the current trial. In error trials a letter 'X' appeared above the FC to indicate time-over and a number '0' (zero) appeared below the FC to indicate no points were acquired.

3.2.4. Experimental Design

Three different cursors were used in this study. Each cursor had its acute angle turned to a different direction (90º, 180º, 270º) and was assigned to one of three different key maps. In this study, we used three distinct key-maps (KM: KM1, KM2, KM3, Figure 2B), which were indicated by different shapes of the cursors (triangles with different spatial acute angle orientation). A KM was associated to a three-move directions rule which subjects had to learn by trial-and-error.

An implicit KM rule associated the index finger with the motion direction assigned to the acute angle of each cursor. Also, five differing start-goal pairs (SG1, SG2, SG3, SG4, and SG5, Figure 2C) were prepared and arbitrarily combined to one of the three KMs to form KM-SG sets. The optimal sequence length varied for each KM-SG set, though we tried to set the optimal length to six moves long. In general, there were several OAS for a single KM-SG set. However, in order to avoid performance variability and to observe learning-related performance improvements, we asked subjects to learn a single OAS for each KM-SG set, and to perform as fast and accurately as possible.

We had subjects perform one training session and one test session in two consecutive days with one night sleep interval between them. Both training and test sessions were further divided into two trial blocks with a 5-minute rest interval. In the training session four different KM-SG sets (KM1-SG1, KM1-SG4, KM2-SG2, KM2-SG5, Figure 2D) were chosen for learning, and each was practiced for a total of 40 trials (20 trials per training block: 10 immediate start and 10 delayed start trials).

In the test session nine different combinations of KM-SG1 sets were used, which consequently defined the three task conditions depending on the familiarity in the training session: *Condition I:* a new key-map (KM3) was introduced so that subjects had to learn both the new KM rule and appropriate action sequences from scratch. *Condition II:* the key-maps used in the training session (KM1, KM2) were combined with new pairs of SG positions to form new KM-SG sets (KM1-SG2, KM1-SG3, KM2-SG1, KM2-SG3) so that subjects had to only learn new action sequences. *Condition III:* KM-SG sets used in the training session (KM1-SG1, KM2-SG2) were retested and subjects

had to execute the previously well-learned action sequences. A single KM-SG set was practiced for a total 20 trials (10 trials per test block: 5 immediate start and 5 delay start trials).

These task conditions were intended to mimic and to separately asses the three different stages of motor sequence learning depicted in Figure 1 during the same scanning session. Each trial block consisted of repetition of two rounds of trials with different SG-KM sets (four in training and nine in testing sessions), first in the immediate response start condition and then in the delay response start condition. The order of SG-KM sets was randomized in each round of trials.

3.2.5 Behavioral Analysis

The following behavioral performance variables were chosen for analysis: a) reward score – number of points acquired for each trial; b) number of moves (NM) – number of key-presses executed to reach the target goal; c) reaction time (RT) – time from go signal to execution of first key-press; d) execution time (ET) – time from first key-press to execution of last key-press. Successful and error trials were included in the analysis of the reward variable, but only successful trials were analyzed for the other performance measures.

A series of Wilcoxon signed rank tests was performed on the four behavioral performance measures in the training session to test for learning-related improvements

within and between practice blocks and to test the effect of start time condition on behavioral performance. The performance variables (reward score, number of moves, RT and ET) in the test session were subjected separately to a series of 3 x 2 x 2 (Test Conditions x Start Time Condition x Test Block) multi-way analyses of variance (ANOVA) with repeated measures, and an additional series of *multiple comparison procedures* and Tukey's post hoc comparisons to find for significant differences between pairs of levels within a factor. A one-way ANOVA was carried out to identify possible effects of KM-SG set on subjects' performance. The statistical analysis was performed in MATLAB (version 7.0) using statistical toolbox.

3.3. Results

3.3.1. Behavioral Results in the Training Session

Figure 3 summarizes the time course of the subjects' average performance during the training session. The four behavioral performance measures in the immediate (gray lines) and delay start (black lines) conditions are shown for the two KM-SG sets retested in the test session (KM1-SG1, KM2-SG2). The average reward score increased gradually (Figure 3A) as the number of movements to reach the goal position decreased toward the optimum of six key-presses (Figure 3B). The average reaction time was significantly shorter ($p<.001$) in the delayed start condition than in the immediate start condition (Figure 3C).

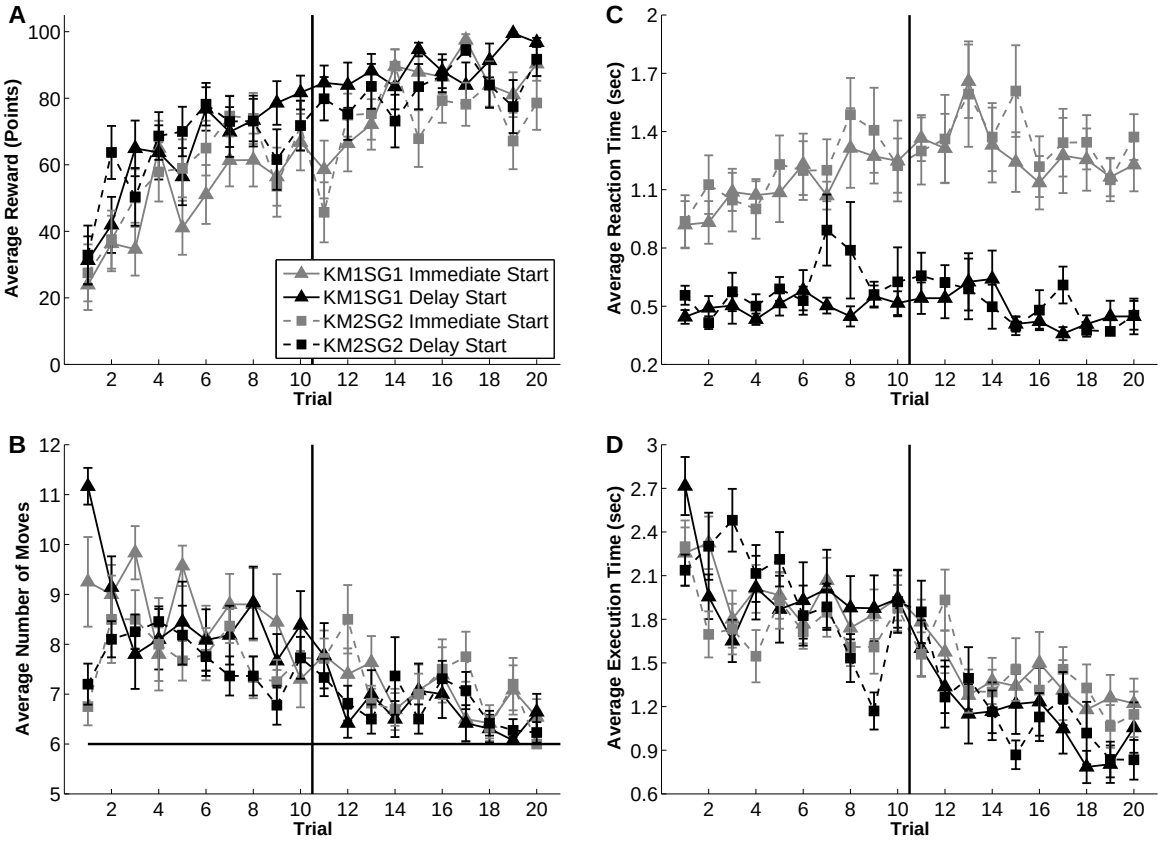

Figure 3. Trial-by-trial time course of performance improvement in the training session. (A) Reward score. Only the performance of the KM-SG sets retested in the test session is displayed in this figure. The vertical black line divides the performance of training blocks 1 and block 2. Error bars represent standard errors.
(B) Number of moves to reach the target goal. The horizontal line represents the length (six moves long) of the optimal sequence for each KM-SG set.
(C) Reaction Time. This is the interval time between the presentation of the go signal and the onset of the first key-press during the execution period.
(D) Execution time. This is a measure of the time interval between the onset of the first and last key-presses.

There was an intriguing increase in the reaction time under the immediate start condition in the middle of the training session. We compared the reaction time in three phases arbitrarily defined: initial (trial 1-7), intermediate (trial 8-13) and late (trial 14-20), and

found that the reaction time in the intermediate phase was significantly longer than in the initial phase ($p<.05$, Wilcoxon signed-rank test). This suggests that subjects tended to think more before acting in the immediate start condition in the middle stage of learning. The execution time, measured as the time interval from onset of the first key press to the last key press, also followed a gradual decrease as learning took place, implying the formation of sequential action chunks (Figure 3D).

3.3.2. Behavioral Results in the Test Session

3.3.2.1. Optimal Action Sequence Pathways

Figure 4 shows examples action sequences learned by two representative subjects. The task could be solved by multiple optimal pathways leading to the goal position. However, subjects were instructed to learn one single optimal action sequence for each KM-SG set. The action sequences were calculated as the ones with the same order of finger key-presses (e.g. ring, middle, middle, middle, index, ring) and which subjects performed with the highest frequency throughout the experiment. As can be observed in Figure 4, subjects learned different optimal action sequences for the same KM-SG set, but in some cases they learned the same action sequence.

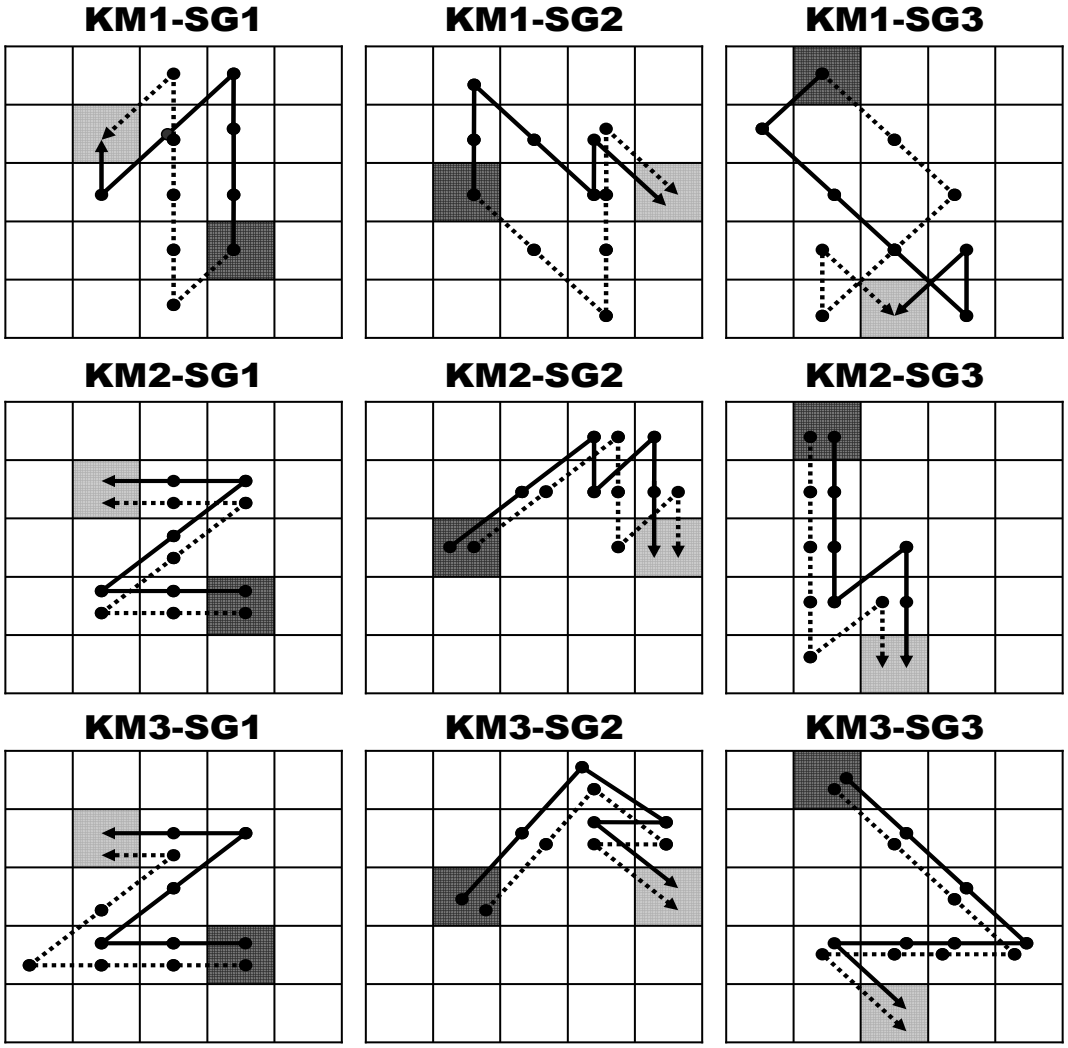

Figure 4. Examples of cursor movement paths learned by two representative subjects for all KM-SG sets used in the test session. Subject 06 (dotted lines), and subject 12 (continuous lines).

3.3.2.2. Reward Score

Figure 5 (top row) shows the trial-by-trial time course of the reward score averaged over all subjects in each task condition. In Condition I, although a new key-map (KM3) was introduced, the performance was unexpectedly better than that found in the training

session (see Figure 3A). These results suggest that there was considerable generalization while using a new key-map, probably due to problem solving skills acquired during the training session, when the task was completely new. In Condition II, the use of pre-learned key-maps for new start-goal positions, subjects performed better under the delay start, although the time-course of the reward score was surprisingly lower than that of Condition I, the reason for which will be discussed below. In Condition III, the retest of the pre-learned KM-SG sets, the performance was almost highly accurate. Significant statistical effects were revealed for Test Conditions, $F(2,2877) = 83.58$, $p<.0001$, Start Time Condition, $F(1,2878) = 25.8$, $p<.0001$, and Test Block, $F(1,2878) = 36.87$, $p<.0001$. A significant two-way interaction also emerged for Test Condition x Start Time Condition, $F(2,2877) = 3.89$, $p =.02$, Test Condition x Test Block, $F(2,2877) = 7.61$, $p <.001$, and Start Time Condition x Test Block, $F(1,2878) = 5.13$, $p=.02$. No significant three-way interactions were observed.

We investigated the extent of the beneficial effect of delay start by computing the reward gain, defined as the difference in the reward score between delay and immediate start trials. The results demonstrated that there the reward gain was significantly greater in Condition II than in the other two conditions (one-way ANOVA, $p < .001$, Figure 6).

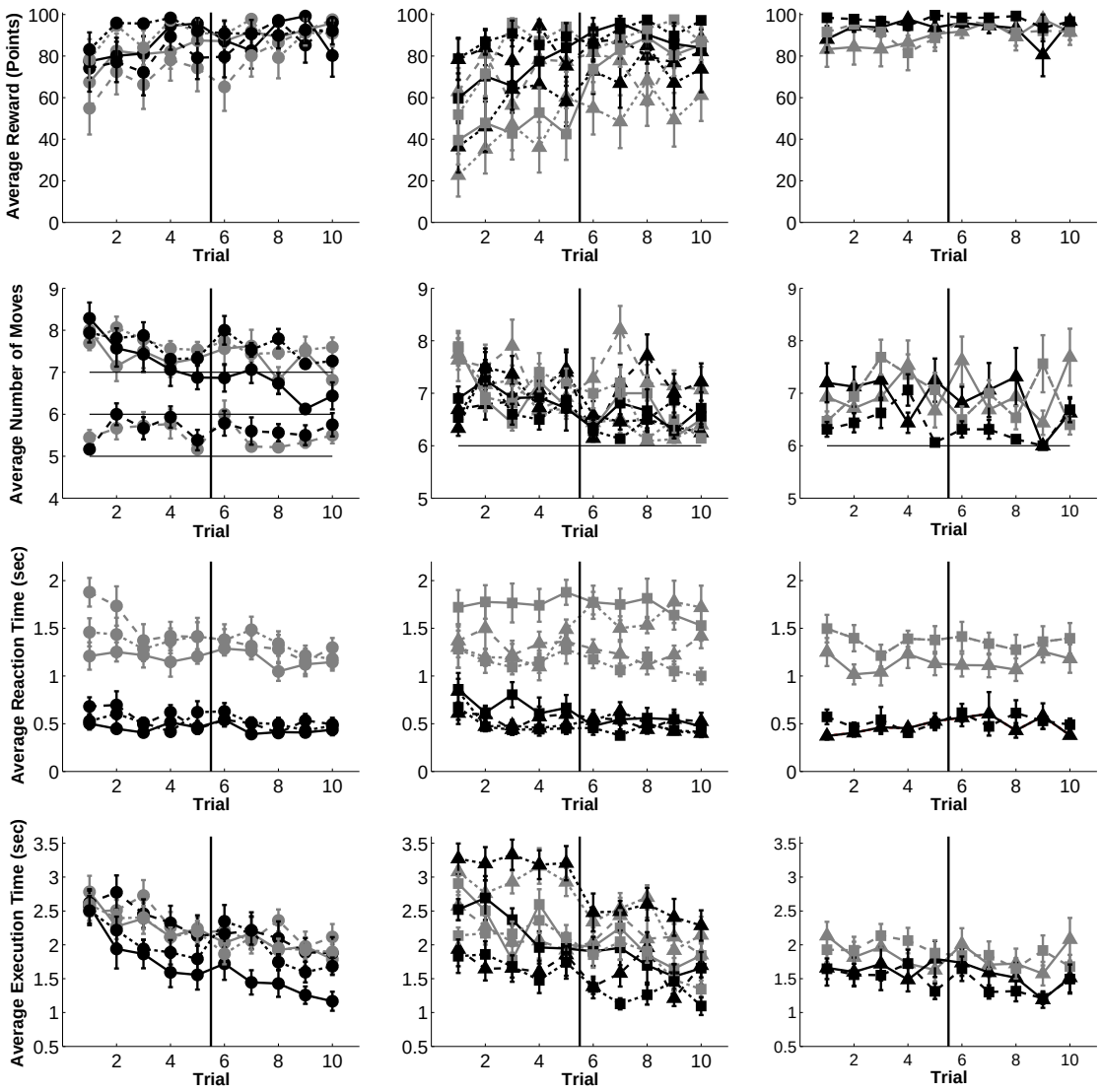

Figure 5. Trial-by-trial time course of average performance for the task Conditions I, II, and III in the test session.
(A) Reward score.
(B) Number of moves to reach the target goal.
(C) Reaction time. Time interval between the go signal and the onset of first key-press in the execution period.
(D) Execution time. Time interval between the onset of the first and last key-presses in the execution period. Error bars represent standard errors. The vertical black line divides the performance of test blocks 1 and 2. The horizontal black in B represents the length of the optimal actions sequences

Tukey's Post hoc comparisons of the means revealed that subjects' performed was higher when a response started after a delay period (M = 87.4, SE = 1.9) than when it had to be started immediately (M = 80.82, SE = 4.7). In addition, the ANOVA estimated coefficients showed that the reward gain under the delay start condition was larger for Test Condition II (c: 2.3 points) than for Test Condition I (c: -1.3 points) and Test Condition III (c: -1 points). A one-way ANOVA showed that there was significant difference in the performance among the different KM-SG sets $F(8,2871) = 42.76$, $p<.001$, revealing that the sets in Condition 1 had the highest reward score, and that the KM1-SG3 set (M = 56, SE = 5.53) in Condition II had the lowest group performance, suggesting that some intrinsic level of difficulty for action sequence learning was associated with certain KM-SG sets.

3.3.2.3. Number of Moves

Figure 5 (second row) shows the number of moves in successful trials. Note that in Condition I, the optimal number of moves were 6 (KM3-SG1), 5 (KM3-SG2), and 7 (KM3-SG3). No significant effect was observed for Test Conditions. There was a significant effect of start time condition, $F(1,2473) = 8.2$, $p < .005$, and Test Blocks, $F(1,2473) = 14.39$, $p<.001$. No two-way or three-way ANOVA interactions were observed.

3.3.2.4. Reaction Time

Figure 5 (third row) shows the reaction time, from go signal to onset of the first key press. Subjects responded faster under the delayed start condition, $F(1,2462) = 1828.53$, $p<.001$. An overall faster RT was also observed for Task Condition III, $F(1,2462) = 6.32$, $p<.001$. The one-way ANOVA revealed that subjects took longer to start a response for KM2-SG1 set ($M = 1.1$ secs, $SE = 0.9$) in Condition II.

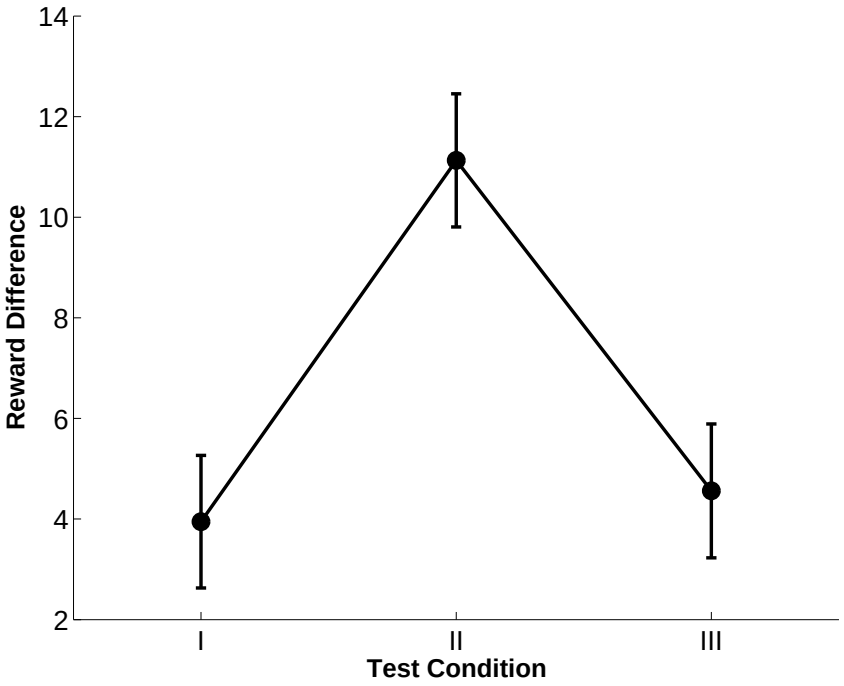

Figure 6. Reward gain computed as the difference, collapsed over the two trial blocks, of average reward in delayed and immediate start trials for the three task Conditions. Error bars represent standard errors.

3.3.2.5. Execution Time

The execution time was significantly affected by the Task Conditions, $F(2,2445) = 41.94$, $p < .001$ and Start Time Condition, $F(1,2446) = 56.35$, $p < .001$ (Figure 5, bottom row). The ET significantly decreased between Test Blocks, $F(1,2446) = 63.64$, $p < .001$, as learning went on. A significant two-way interaction between Test Condition x Test Block was found, $F(2,2445) = 4.32$, $p < .05$. Among the nine KM-SG sets, KM1-SG3 (in Condition II) had significantly longer ET ($M = 2.71$, $SE = 0.75$).

3.4. Discussion

In light of Reinforcement Learning (RL) theory, we predicted that subjects would utilize distinct RL action selection strategies based on the degree of experience and the time available for control. In the very early stage of learning a new task, subjects would use a model-free RL strategy, such as the action value-based learning. As knowledge of the task setting is acquired, the use of a model-based RL strategy would be the most appropriate, such as planning by the forward model, provided that there is enough time for the increased cognitive load. After extensive training, subjects would utilize a model-free, memory-based strategy to reproduce pre-learned action sequences.

In order to test this hypothesis and investigate the neural implementation of the computational processes, we developed a new behavioral paradigm called ‘grid-sailing task’, in which subjects were requested to execute sequential finger movements to move a cursor from its start position to a goal position. In order to simultaneously reproduce distinct learning stages during a single experimental session, subjects were requested to perform the task under three test conditions: Condition I: learning new action sequences with a new key-map; Condition II: use of learned key-maps to learn new action sequences, and Condition III: reproduce well-learned action sequences with well-learned key-maps. The time available for planning was manipulated by introducing two start time conditions, immediate or delay.

In the training session with two fixed KM-SG sets, the subjects’ performance in the reward, the number of moves, and the execution time improved gradually (Figure 3 A, B, D). On the other hand, we found a significant increase in the reaction time in the immediate start condition in the intermediate stage of learning, which could be the result of the usage of model-based planning by the forward model learned up to this stage.

In the test session, subjects’ performance in Condition I using the newly introduced KM3 (Figure 5, left column) was unexpectedly better than that in the training session. This suggests that there was a considerable effect of generalization even with the use of a new key-map. This would include the general acquaintance with the task, such as any heuristics in finding a right path. Furthermore, we had adopted an implicit rule in the key-maps such that the index finger key press resulted in a cursor movement to the direction of the acute angle of the triangle and the middle and ring finger key presses

corresponded to the movements in the clock-wise directions (see Figure 2B). This constraint was introduced to make the task easy to learn during a limited duration of the test session, but in fact it may have facilitated learning of the newly introduced key-map. In particular, the KM3 was a simple 90° anti-clock wise rotation of KM1, which may have allowed subjects to quickly acquire the state transition model for the new KM3.

The performance in the Condition II, when subjects used previously learned key-maps, was unexpectedly worse than in the Condition I. A possible reasons for this intriguing result is that the particular combination of KM1-SG3 in Condition II required complex paths with self-crossing, with subjects' performances were especially worse in this setting. Counter-balancing of distribution of different KM-SG sets to different task conditions among subjects are desired in the future experiments. Despite this deviation from our prior expectation, we observed a significantly improved performance in delayed start condition compared to immediate start condition in the Condition II than in Conditions I and III (Figure 6). Subjects could benefit from the additional time from the task presentation to the start of the movement especially in Condition II. The most likely reason for this is that the subjects utilized the pre-learned state transition models of KM1 and KM2 for planning of multi-steps paths to reach a new goal position in Condition II. In Condition I, with newly introduced KM3, subjects could not plan any particular path, at least in the early stage of learning. In Condition III, on the other hand, the optimal paths were already memorized so that a good performance could be achieved without a delay start.

The idea of parallel learning systems has received much attention recently and several models have been proposed, for example, systems that learn by using visual- and body-based coordinates (Hikosaka et al., 1999), a fast-learning-fast-forgetting and slow-learning-robust-memory systems (Smith, Ghazizadeh and Shadmehr, 2006; Lee and Schweighofer, 2009), implicit-explicit or declarative-procedural learning systems (Keele et al., 2003). An important issue is how the brain arbitrates among several control systems to guide behavior, although a few models have been proposed (Bapi and Doya, 2001; Daw et al., 2005; Shah and Barto, 2009).

3.5. Conclusion

Based on RL theory, we proposed that parallel and distinct action selection strategies, Model-free RL, Model-based RL, and recall of learned policy, are predominantly implemented at each learning stage. In order to test this hypothesis, we developed a novel grid-sailing task, which is simple but calls for non-trivial planning of action sequences. Our finding of the improved performance under delayed start condition particularly when the subjects used pre-trained key-maps supports our hypothesis that humans use model-based strategy if they are given enough time for planning. The finding that the subjects did not gain from the additional time after extended training also supports the hypothesis that subjects use simple recall of a learned policy. As a next step, we are testing which RL learning algorithms, including model-free and model-based, fit

the subjects' choice sequence data to further clarify what strategy subjects used at which learning stage, and the brain mechanisms that implement such strategies.

CHAPTER 4

Multiple Prediction Models for Action Selection in Prefrontal-Basal Ganglia and Cerebellar Networks

4.1. Introduction

4.1.1. Action Learning in Motor Learning

Humans can learn to select actions from scratch, by trial-and-error, or by using knowledge from past experiences. After repeated practice, action selection may rely on predictive mechanisms, such as retrieval of motor memories from well established sensory-motor mappings, or the construction and use of internal models of the dynamics of newly encountered environments.

In classical theories of motor behavior, there is a dominant view that action learning manifests itself in a stage-wise manner (Ackerman, 1988; Adams, 1971; James, 1914) through the operation of a serial transition mechanism. Fitts and Posner (1967), for example, have proposed a three-stage action learning model: an early cognitive/attentional stage where actions are selected on a quasi-random basis to explore

and understand the environment-action relationships; an intermediate associative stage characterized by action planning and possibly prediction of action outcome, and an autonomous stage when habits are formed and can readily be performed.

4.1.2. Classical View of Brain Mechanisms in Motor Learning

Regarding the brain mechanisms, the classical view was of a simple transition of control from cortical to subcortical structures (Sherrington, 1906; Ashby, Turner and Horvitz, 2010). However, although such behavioral changes are well accepted, this classical stage-wise thinking cannot accommodate the recent increasing number of computational (Nakahara et al, 2001), neuroanatomical (Alexander et al, 1986; Haber, 2003), behavioral (Jueptner et al, 1997; Hikosaka et al, 1995) and electrophysiological and imaging studies (Hikosaka et al, 1999) providing evidence that humans and animals might utilize parallel instead of serial mechanisms for action selection and that the brain contains multiple parallel systems specialized for different types of information processing.

4.1.3. Working Hypothesis

The working hypothesis in this thesis is that the brain contains multiple parallel controllers for action selection, with different state and action representations and whose dominant role is a function of an individual experience and context-dependent; these controllers simultaneously implement different computations, by distinct action selection

algorithms, on the same or different input and generate multiple output signals that might interact to produce optimal action selection.

In our previous work (Fermin et al, 2010) we assumed a mechanism for selecting and combining the outputs of the multiple learning algorithms depending their learning experience by multiplicative integration of action selection probabilities. To test this hypothesis we borrowed concepts of *transfer of learning* (Shea and Morgan, 1979; Duncan, 1953; Barnett and Ceci, 2002; Doane et al., 1995; Woodward, 1943) in concert with Reinforcement Learning (RL, Taylor, 2009), a computational theory of adaptive optimal control with strong background on biological and psychological processes (Sutton and Barto, 1998).

4.1.4. Reinforcement Learning and Action Selection

RL has played an important role in understanding the processes of action selection and learning in humans (Sutton & Barto, 1998) and other animals such as rats (Daw, et al, 2005), monkeys (Hikosaka et al, 2002), and birds (Doya and Sejnowski, 1999). RL proposes two methods for action selection that, model-free and model-based, resemble the way humans learn.

The model-free method uses action value functions and a policy, learned by trial-and-error from actual exploratory experiences, to predict future rewards based on current sensory input states and available actions. In situations where the environment changes,

the policy must be updated based on the difference of the predicted and actual rewards. The model-based method employs an internal forward model, assuming that it has been previously acquired, to predict the future state reached by a hypothetical discrete action or multi-step actions, and uses the results of this mental simulation to evaluate states and actions, and to selecting an action for execution based on its goodness. A model-based method is helpful not only for on-line action planning, but also for off-line learning such as in mental training, by enabling action learning and problem solving even in actual time-limited experience (Doya, 2007). Success of using a model-based method depends, however, on the model accuracy, working memory capacity, time available for computation and the action search depth.

4.1.5. Transfer of Learning and Action Selection

Transfer of learning is concerned with the application and exploitation of prior experience (e.g. motor skills, abstract knowledge) acquired in one situation to another learning or problem solving situation (Lee & Magill, 1983). An essential characteristic of transfer is its capability to facilitate and boost the speed of learning a new task. In transfer of learning an initial bias such as the reuse of learned action values, state values or a reward function (Taylor, 2009) is provided, although an update of these priors might be necessary. Consequently, transfer of learning bypasses an initial learning stage for there is no need to re-learn the common features shared by two tasks, a process called savings.

4.1.6. Testing Reinforcement Learning and Transfer of Learning

How can we conciliate concepts of transfer of learning and the RL algorithms with the stage-wise manifestation of action selection learning and our hypothesis of parallel action selection systems?

We designed a task paradigm by which we should be able to break apart the different learning stages, and accordingly, assess the performance of the multiple parallel action selection systems separately, and the brain mechanisms involved in their implementation. For that purpose, we had human subjects perform a grid-sailing task, inside the fMRI scanner, under three task conditions that mimicked different levels of experience in action selection such as action selection learning from scratch (condition 1), by transfer of learning (condition 2), and selection of habitual actions (condition 3), each task was theoretically equivalent to one of three action learning stages as proposed by Fitts and Posner (1969) and more recently by Hikosaka et al. (1999).

In the grid-sailing task, subjects had to move a cursor, associated to a three-move direction key mapping rule, from its start position to a goal position by performing a sequence of finger movements. In condition 1, subjects had to learn new action sequences with an unknown key mapping; in condition 2, subjects used learned key mappings to learn new action sequences, and in condition 3 subjects performed well-learned action sequences.

4.1.7. General Hypothesis for Action Selection and Brain Mechanisms

We have previously hypothesized that humans implement different computations and algorithms depending on their level of experience with a given task and environment (Fermin et al., 2010; Doya, 1999). In the absence of an internal model able to predict an action outcome (condition 1), a model-free action value-based algorithm is likely to take the dominant role in the initial stage of learning by increasing the value of state-action pairs that maximize the cumulative reward. As learning goes on and an internal model of the task is established (condition 2), a model-based algorithm would take over control and play a dominant role in directing action generation regardless of the computational cost and need of more time for internal simulation. As learning proceeds to a late stage after repeated practice and the environment becomes fully predictable (condition 3), a major role in action selection would be played by the learned policy of model-free algorithm as it requires less computation in real time, allowing prompt preparation and efficient execution of an action response. Figure 1 shows the main behavioral characteristics as learning starts from scratch, the RL computational strategies that might be used in different action learning stages and the expected behavioral performances under different task manipulations. The details of these task conditions are described in the following sections.

We also sought for the neural mechanisms likely to implement model-free and model-based methods. It has been proposed that the dorsolateral striatum implements a model-free strategy, and the dorsolateral prefrontal cortex implements the model-based method (Daw et al, 2005). On the other hand, Doya (1999) has proposed that such methods can

be implemented by a combination of output signals of parallel circuit networks linking the cerebral cortex, the basal ganglia and cerebellum.

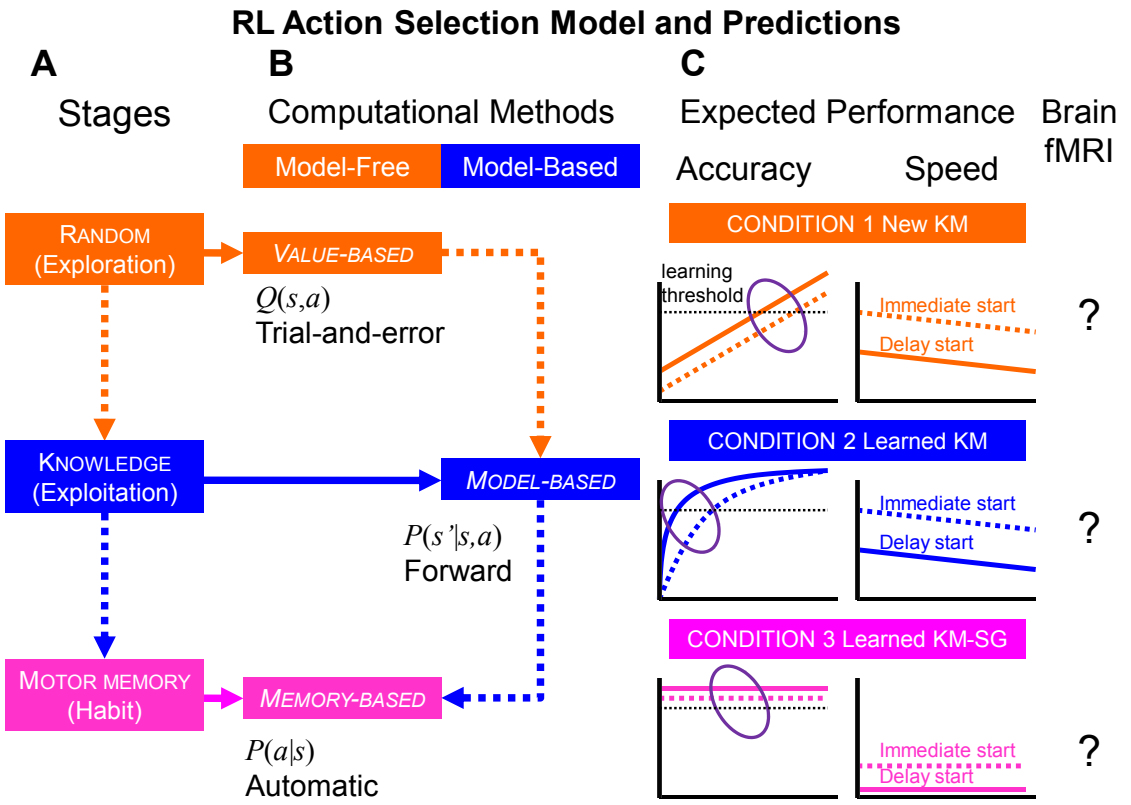

Figure 1. Schematic representation of RL action selection strategies and predicted performance under different task manipulations.
(A) Main behavioral learning characteristics according to RL.
(B) RL algorithms for action selection and learning. The horizontal continuous line arrows represent the parallelism of the systems. The broken arrows represent the different the stage transition in learning and respective action selection methods.
(C) Expected performance under the control of each learning system. The colored dotted lines represent the expected performance under the immediate start condition and the continuous line the performance starting after the delay period. As a measure of accuracy, an 80% average reward score was taken as the learning threshold. Therefore, we expected a faster reaching of the learning threshold in Condition 2 under the delay start trials than in Condition 1. For the speed performance, a faster reaction time was expected for trials starting after the delay period than in trials with an immediate start. In addition, reaction time should be shorter in Condition 3 under both star time conditions than the reaction time in Condition 1 and Condition 2. An fMRI experiment was carried out to identify the brain areas predominantly activated while subjects performed each of the three task conditions.

4.2. Materials and Methods

4.2.1. Participants

Eighteen healthy right-handed dominant subjects (4 female; age 21-39, 26 ± 5 yrs), non-musicians, with normal or corrected to normal (by contact lenses) vision participated in this study. All subjects were screened for a history of psychiatric, neurological problems, or drug use at the time of the experiment. Subjects gave written informed consent and the experiment was approved by the ethics and safety committee of the Advanced Telecommunications Research Institute International (ATR). A fixed payment was made to subjects for their participation in this study. Due to technical fMRI pre-processing problems, we report the imaging results of only 16 subjects.

4.2.2. Apparatus

In the training session, performed outside the fMRI scanner, subjects comfortably sat at a table approximately 40-50 cm from the computer monitor. A regular desktop computer keyboard was used for finger key-pressing. Subjects used the index, middle, and ring fingers of the right hand to press the ‘v’, ‘b’, and ‘n’ keyboard buttons, respectively. Stimulus generation and behavioral data recording were done by custom-made programs written in MATLAB (version 7.0), using Psychophysics Toolbox functions (PTB-3). The test session was performed inside the fMRI scanner and the task was displayed through

an imaging projector on a mirror inside the scanner mounted on top of the head coil. A fMRI compatible response optic button box was used obeying the same finger-key association used in the training session.

4.2.3. Behavioral Task Paradigm

We designed a "grid-sailing task" whose goal was to move a cursor from its start position to a target goal position by performing a sequence of finger movements through the shortest pathway. Subjects were given oral instruction of the task rules before the training session began, and received a short verbal feedback orientation during the first training trials to assure they had fully understood how to play the task.

The task trial started with the computer displaying for 3~5 seconds a 5x5 grid (10 cm x 10 cm) environment with a red fixation cross (FC) on top of the center square matching the monitor center position (Figure 1A). This initial trial event corresponded to the inter-trial interval (ITI) period. Next, a pair of start-goal (SG) positions was simultaneously displayed together with a cursor (black triangle) which appeared on top of the start position.

The color of the start position specified the time to start a response: immediate start (green color) – subjects were instructed to start responding immediately; delay start (red color) – subjects had to withhold a response during a delay of 4~6 seconds, and wait for the go signal indicated by the switching of color of the start position from red to green.

The color of all target goal positions was always blue. After the go signal subjects had a maximum response time of 6 seconds, of which they were not explicitly informed. During the response period, a sequence of finger key presses had to be performed to move the cursor from its initial start position to the goal position. Immediate visual feedback of cursor motion was provided after a button had been pressed. Any key-press leading the cursor to cross out the grid boundaries was not allowed and was indicated by a rapid blink of the cursor which remained at the same position.

A single trial had three possible endings depending on the subject's performance: a) successful trial with the execution of an optimal action sequence (shortest path to goal); b) successful trial with the execution of a suboptimal action sequence (longer path); c) error trial with a failure to reach a goal position due to response period time over. Subjects received performance feedback, displayed for 2 seconds at the trial end, and contained information with the number of key-presses, displayed on top of the square above the FC and the reward score acquired at the current trial, and displayed on top of the square below the FC.

A point system was used as reward for teaching subjects an optimal action sequence: 100 points for the execution of an optimal action sequence; a discount of -5 (minus 5) points for each excessive key-press when a suboptimal action sequence was executed and 0 (zero) point in error trials. Error trials were signaled by a letter 'X' displayed above the FC to indicate time over.

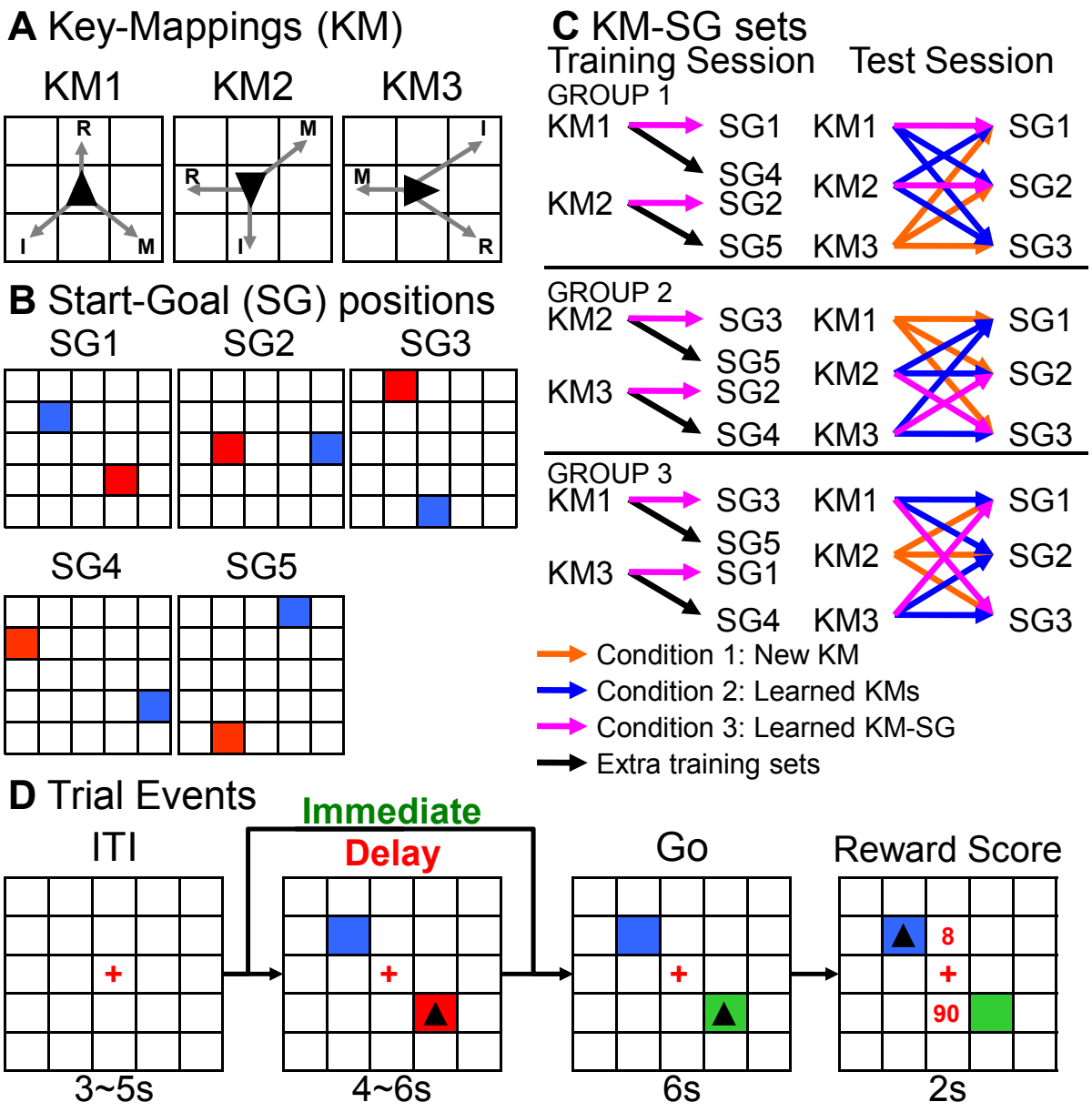

Figure 2. Grid-Sailing Task and Experiment Design.

(A) Key-mappings (KM) and respective fingers (I: index; M: middle; R: ring). The arrows indicate the three directions of motion depending on the current position of the cursor.

(B) Pairs of start-goal (SG) positions. Start position (red) and goal position (blue).

(C) KM-SG sets assigned for practice in the training and test sessions for each group of subjects. Colored arrows in the training session: magenta – KM-SG sets chosen for retest in the test session, black – extra KM-SG sets for training. Colored arrows in the test session: orange – KM-SG sets in Condition 1, blue – KM-SG sets in Condition 2, magenta – KM-SG sets in Condition 3.

(D) Sequence of task flow. The task started with the presentation of a 5x5 grid and a fixation cross (FC) for 3-5s. Next, a pair of SG positions was displayed with a cursor

(triangle), associated to one KM, on top of the start position. When the color of the start position was red it indicated a delay period of 4-6s and the color switching to green indicated the go signal and start of the response period which lasted for 6s. Subjects' task was to execute a sequence of finger movements and to reach the goal by finding the shortest pathway. Performance feedback (2s) containing the number of moves (above FC) and reward score (below FC) was provided at the trial end.

4.2.4. Task Design

Three cursors, all triangle shaped, were used in this experiment, each had its acute angle turned into a spatial direction (90º, 180º, 270º) and was assigned to one of three different key-maps (KM, Figure 1B),which were associated to a three-move direction rule. Also, five differing fixed pairs of start-goal positions (Figure 1C) were prepared and arbitrarily combined to one of the three KMs to form KM-SG sets used in the training and test sessions (Figure 1D).

Subjects started off the task not know neither KMs nor action sequences and had to learn by trial-and-error. A target goal position could be reached by multiple optimal action sequences. However, in order to measure learning-related behavioral changes, subjects were instructed to learn a single optimal action sequence for each KM-SG set, and to perform as fast and accurately as possible. No constraints were imposed on a sequence to be learned, and subjects were free to learn the action sequence they felt most comfortable performing.

We had subjects perform one training session and one test session on different and consecutive days, with one night sleep interval between sessions. Training and test

sessions were further divided into three and two trial blocks, respectively, separated by a 5 minute rest interval. Subjects were randomly assigned to one of three groups and remained in the same group in the two experimental sessions.

For each group in the training session we arbitrarily selected two KMs and each was combined with two different SG pairs. Therefore, subjects had to learn from scratch the two KMs and four different action sequences, one for each KM-SG set. A KM-SG set was practiced for 60 trials (20 trials per training block: 10 immediate start and 10 delay start trials). Each of the four KM-SG sets was performed once in a block of four randomly generated trials with each block alternating between immediate and delay start, and the first block of four trials being immediate start.

In the test session subjects performed the grid-sailing task, inside the fMRI scanner, under three task conditions: Condition 1 – new KM: the KM not used in the training session was introduced and combined with three SG pairs, so that subjects had to learn both the new KM and an action sequence for each newly KM-SG set. Condition 2 – learned KMs: the KMs learned in the training session were combined with different SG pairs, and subjects essentially had to retrieve the appropriate KM rule and use it for planning and learning of new action sequences. Condition 3 – learned KM-SG sets: two of the four KM-SG sets practiced in the training session were chosen for retest in the test session.

A single KM-SG set was practiced for a total 20 trials (10 trials per trial block: 5 immediate start and 5 delay start trials). In total, subjects performed nine different KM-

SG sets in the training session: three KM-SG sets in task condition 1, four KM-SG sets in task condition 2, and two KM-SG sets in task condition 3. Similar to the training session, the task was performed in blocks of nine randomly generated trials, with alternation of the start time every nine trials, and the first block of nine trials set to immediate start. The task conditions were intended to mimic distinct levels of experience in action selection which might particularly depend on the implementation of specific action selection methods.

Subjects performed a short version (6 immediate and 6 delay start trials) of the same task practiced on the previous day in the training session right before the start of the fMRI scanning in order to assure they still remembered the task procedures, and more importantly, the action sequences learned for each KM-SG set. At the end of this short pre-fMRI practice, subjects received explicit instruction regarding the introduction of different task conditions that they had to perform inside the fMRI scanner.

4.2.5. Analysis of Behavioral Data

We analyzed the following behavioral performance variables: a) reward score – number of points acquired in a trial; b) number of moves – number of key-presses executed to reach the target goal; c) reaction time (RT) – time elapse from go signal to onset of the first key-press; d) execution time (ET) – time from onset of the first key-press to the execution of the last key-press. The behavioral measures in the training and test sessions were separately subjected to a series of n-way ANOVA with repeated measures

including the factors: Group, Task Condition, Trial Block, and Start Time Condition. The statistical analysis also included *multiple comparison procedures* and Tukey's post hoc tests to seek for significant differences between pairs of levels within a factor.

4.2.6. fMRI Data Acquisition

Scanning took place at the Brain Activity Imaging Center, Kyoto. Detailed anatomical data were collected using a multiplanar rapidly acquired gradient echo (MP-RAGE) sequence. T2*-weighted echo planar images (EPIs) with blood oxygen level dependent (BOLD) contrast were acquired on a Siemens Tesla 3 Magnet Trio MRI scanner (TR: 2000 ms, TE: 23 ms, FOV: 192 mm, flip angle: 80°). Thirty coronal oblique slices (3 x 3 x 5 mm) were acquired parallel to the plane containing the anterior and posterior commissures (AC-PC plane) after prescribing slice position based on automatic measurements of rotation, translation and tilt of the structural images relative to an average image. The fMRI session was split into two runs, each lasting approximately 22 min, with the number of volumes depending on the subjects' performance.

4.2.7. Analysis of the fMRI Data

Statistical parametric mapping (SPM) with SPM8 software (Wellcome Trust Centre for Neuroimaging, UCL) was used to preprocess all fMRI data, which included correction for slice time acquisition, spatial realignment of all volumes to the first image, spatial

normalization using regressor parameters estimated by the unified segmentation process with voxel size resampled to 3 x 3 x 3 mm, and smoothed using a Gaussian kernel with an isotropic full width at half maximum of 8 mm. A high-pass temporal filtering with a cutoff of 128 s was applied to remove low-frequency drifts in signal, and global changes were removed by proportional scaling. Statistical analysis was conducted using a general linear model and a set of boxcar functions on three events (ITI, Response Period and Reward delivery) of the immediate start trials and on four events (ITI, Delay, Response Period and Reward delivery) of the delay start trials. Subject-specific design matrices were created using the above trial events for each of the tree task conditions, including a nuisance partition containing the head motion regression parameters estimated in the realignment procedure.

A series of contrast images were generated for each subject by using a subtraction approach and were subsequently taken to a second-level for a random effects group analysis using one-sample T tests. We were primarily interested in the specific effect of task condition on the delay-related period activity, and all areas of interest are reported at $p<0.001$, $p<0.005$, and $p<0.01$ (uncorrected; see results section for details). A region of interest analysis (ROI) was carried out using Marsbar and neuro-anatomically defined mask images generated using the Pickatlas Toolbox and the Anatomy Toolbox. The mask image for the basal ganglia ROI analysis contained the caudate nucleus and putamen, and the graphs were plotted using the Multicolor Toolbox.

4.3. Results

4.3.1. Behavior in the Training Session

4.3.1.1. Reward Score

Figure 2 summarizes the group average performance during the training session, where only the behavior data of the KM-SG sets retested in the test session of all three groups was pooled together for analysis. A three-way ANOVA found a main effect of training group on the reward score, $F(2,2142)=19$, $p < 0.001$, and a post hoc comparison using the Tukey's test indicated that the mean reward score for group 2 ($M = 83.02$, $SE = 2.87$) was significantly different than the group 1 ($M = 73.3$, $SE = 3.27$) and group 3 ($M = 72.98$, $SE = 3.6$), and no significant difference was observed between group 1 and group 3. There was also a main effect of start time on subjects' behavior, $F(1,2142)=33$, $p < 0.001$. This effect showed that a subject could increase a reward score and consequently reach the goal position more effectively when a response started after a delay than when it had to be started immediately.

There were significant performance improvements across trial blocks, $F(2,2142)=216$, $p < 0.001$, suggesting that the total average reward gradually increased (Figure 3A) as subjects learned he KM rule associated to each cursor, and optimal action sequences leading to the goal position. Finally, there was an interaction between training group and start time, $F(2,2142)=4.5$, $p \leq 0.01$, training group and trial block, $F(4,2142)=10.92$, $p \leq$

0.01, start time and trial block, $F(2,2142)=5.72$, $p \leq 0.005$, but no interaction effects were found for training group x start time x trial block at $p < 0.05$. Although group differences were observed in the training session, the analysis of subjects' behavior in the short practice, right before the start of the fMRI session, revealed that these group differences completely disappeared with a one night sleep interval. Similar learning-related improvements in performance were found for the number of moves (figure 3D), reaction time (figure 3E), and execution time (figure 3F).

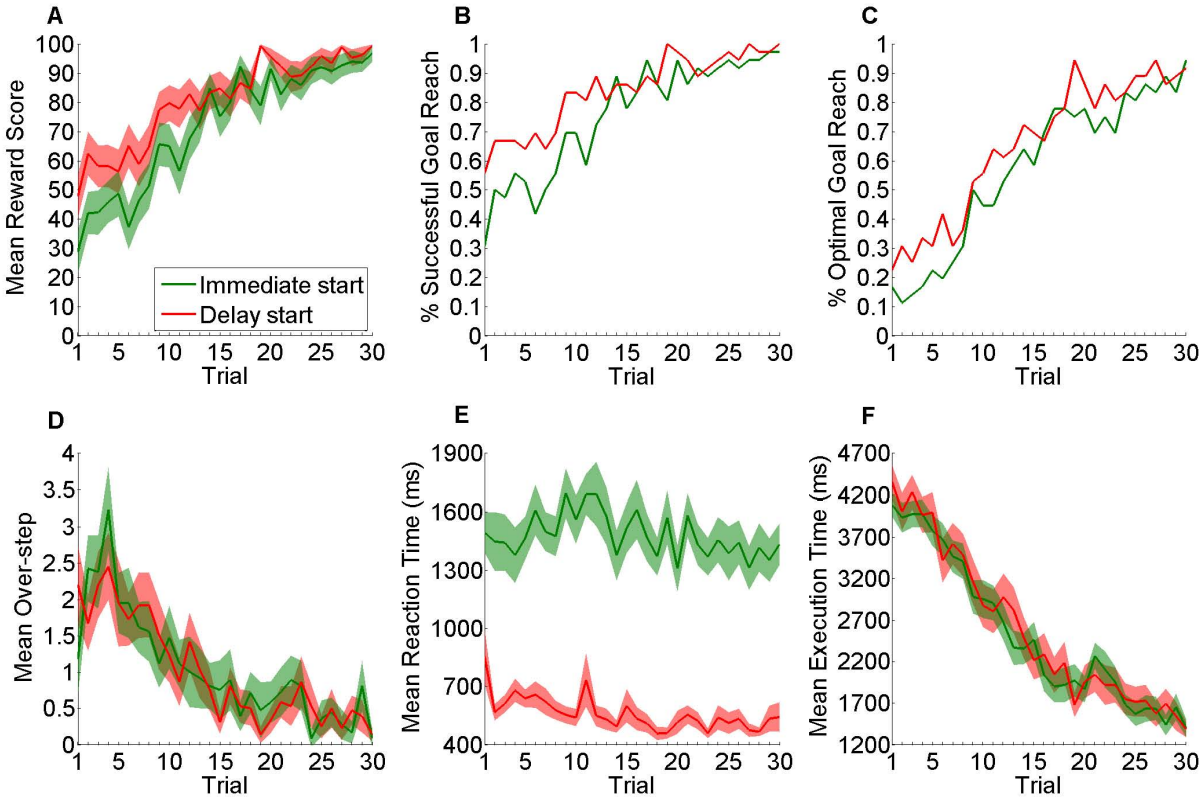

Figure 3. Behavior Performance in the Training Session and Acquisition of Sequential Motor Memories

(A) Mean reward score. The reward score was averaged in blocks of every 5 trials across subjects. The same procedure was applied to the other behavioral measures.

(B) Percentage of trials with successful goal reach. In this analysis optimal and non-optimal goal reach trials were analyzed. All subjects reached the learning criteria of 80% successful performance.

(C) Percentage of trials with optimal goal reach. Only the trials in which subjects performed an optimal action sequence was analyzed.
(D) Mean over-step. This is a measure of the number of moves that exceeded the optimal sequence length.
(E) Mean reaction time (ms). A significant increase in the reaction time was found between trials 6 and 15 ($p < 0.01$) under the immediate start condition.
(F) Mean execution time (ms). No significant difference was found in the performances under the immediate and delay start conditions.

4.3.1.2. Reaction Time

The three-way ANOVA with repeated measures showed a significant effect of training group, $F(2,2139)=69.9$, $p < 0.001$, start time, $F(2,2139)=1451.32$, $p < 0.001$, trial block, $F(2,2139)=6.07$, $p < 0.005$, training group and start time, $F(2,2139)=14.32$, $p < 0.001$, training group and trial block, $F(2,2139)=9.92$, $p < 0.001$, and training group * start time * trial block, $F(2,2139)=3.81$, $p < 0.005$. The post hoc test revealed that the overall reaction time in training group 3 ($M = 1.22$, $SE = 0.09$) was longer than the reaction time of training groups 1 ($M = 0.92$, $SE = 0.07$) and 2 ($M = 0.91$, $SE = 0.02$), but no significant difference between training group 1 and 2. We also found a significant increase in the reaction time under the immediate start condition during the first part of the training session, but which subsequently decreased as learning progressed (Figure 3E).

4.3.1.3. Execution Time

A significant effect was found for training group, $F(2,2139)=57.76$, $p < 0.001$, trial block, $F(2,2139)=471.62$, $p < 0.001$, and a two-way interaction effect for training group and trial block, $F(2,2139)=10.38$ (Figure 3F). The post hoc test revealed that the overall execution time was significantly different among the three training groups with group 1 having longer execution time ($M = 2.92$, $SE = 0.04$) than group 2 ($M = 2.22$, $SE = 0.04$) and group 3($M = 2.53$, $SE = 0.04$).

4.3.1.4. Number of Key-Presses to Reach a Goal

The three-way ANOVA with repeated measures showed a significant effect of training group, $F(2,2142)=84.03$, $p < 0.001$, trial block, $F(2,2142)=91.17$, $p < 0.001$, a two-way interaction effect for training group and start time, $F(2,2142)=5.63$, $p < 0.005$, and training group and trial block, $F(2,2142)=4.37$, $p < 0.001$. No other significant interaction effects were found. The post hoc analysis using the Tukey's test revealed that the number of moves to reach the goal position was significantly larger in group 1 ($M = 7.66$, $SE = 0.07$) than in group 2 ($M = 6.33$, $SE = 0.07$) and group 3 ($M = 6.52$, $SE = 0.07$), but there was no significant difference between groups 2 and 3 (Figure 3D).

4.3.2. Behavior in the Test Session

4.3.2.1. Reward Score

We merged the data of all subjects from the three groups and carried out the statistical analysis with the null hypothesis of no significant differences in the reward score performance among the three task conditions. A three-way ANOVA with repeated measures found a main effect of task condition $F(2,3228)=181.94$, $p < 0.001$, start time $F(1,3228)=121,13$, $p < 0.001$, and trial block $F(1,3228)=126.6$, $p < 0.001$. Significant interaction effects were also found for task condition and star time $F(2,3228)=17.24$, $p < 0.001$, task condition and trial block $F(2,3228)=25.99$ $p < 0.001$, start time and trial block $F(1,3228)=10.81$, $p = 0.001$. A significant three-way interaction for test condition x start time condition x trial block, was also found $F(2,3228)=3.42$, $p = 0.0328$.

Figure 4A shows the trial-by-trial time course of the average reward score for the three task conditions. As a measure of speed of learning an 80% performance accuracy as the average reward score was used. The learning threshold was achieved by subjects in Condition 1 under the delay start time condition only in the 7th trial and performance was still oscillating in the following practice trials. On the other hand, the learning threshold was achieved faster in the 3rd trial by subjects in Condition 2. The learning threshold was reached by subjects in Condition 1 and Condition 2 under the immediate start condition only in the last trial of the task. The reward score performance in Condition 3 remained above the learning threshold throughout the experiment.

A post-hoc analysis for data dredging was carried out using the Tukey's honestly significant difference test to seek for significant differences between pairs of data. This analysis revealed a significant difference in the reward score between Condition 1 ($M = 64.03$, $SE = 1.1$) and Condition 2 ($M = 72.21$, $SE = 0.96$), < 0.001, Condition 1 and Condition 3 ($M = 96.73$, $SE = 1.3$), $p < 0.001$, and Condition 2 and Condition 3, $p < 0.001$. The post hoc analysis revealed a significant increase across trial blocks in the reward score of Condition 1 in immediate start trials, $F(1,538)=48.41$, $p < 0.001$ and delay start trials, $F(1,538)=26.19$, $p < 0.001$. Similar result was found in Condition 2 in immediate start trials, $F(1,718)=91.06$, $p < 0.001$ and delay start trials, $F(1,718)=32.78$, $p < 0.001$. The post-hoc test failed to identify any changes in reward score performance in Condition 3 in immediate start trials, $F(1,354)=0.003$, $p = 0.9995$ and delay start trials, $F(1,354)=0.22$, $p = 0.6417$ (Figure 4B).

We investigated the effect of the immediate and delay response start on the subject's ability to plan and execute action sequences. The post hoc test showed a significant effect of response start time for Condition 1 (immediate start: $M = 55.52$, $SE = 2.63$; delay start: $M = 72.53$, $SE = 2.26$), $p < 0.001$, and for Condition 2 (immediate start: $M = 60.56$, $SE = 2.57$; delay start: $M = 84.02$, $SE = 2.08$), $p < 0.001$. We also found a significant effect of response start time on the reward score of Condition 3 (immediate start: $M = 94$, $SE = 1.22$; delay start: $M = 98$, $SE = 0.85$), $p < 0.001$. No significant difference was found in the reward score between Condition 1 and Condition 2 under the immediate response start, $p = 0.0578$. However a significant difference was found for trials under the delay start response, $p < 0.001$ (Figure 4C).

To further investigate the extent of the beneficial effect of delay response start on improvements in the reward score, we analyzed the reward gain, defined as the difference between the reward score of the delay and the immediate start trials for each of the three task conditions. The reward gain was first computed individually and then taken to a group analysis using a two-way ANOVA with test condition and trial block as factors. This analysis revealed a significant effect of test condition, $F(2,1560)=18.23$, $p < 0.001$, and trial block, $F(1,1560)=20.64$, $p < 0.001$ and a significant two-way interaction, $F(2,1560)=4.36$, $p = 0.0129$. The post hoc analysis also revealed a significantly larger reward gain in Condition 2 than in Condition 1 ($p < 0.05$) and this difference was larger in the first trial block ($p < 0.01$; Figure 4D) .

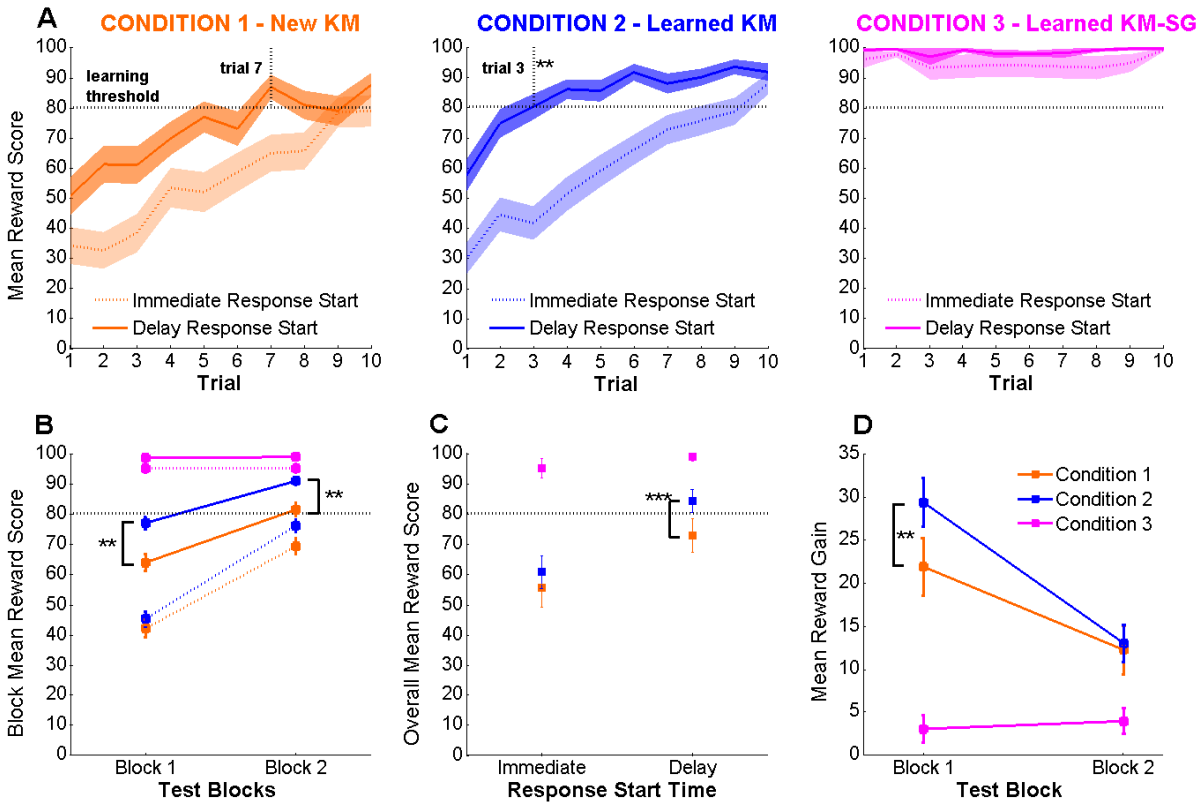

Figure 4. Reward Score Performance in the fMRI Session

(A) Mean reward score. Trial-by-trial time course average of the reward score across subjects. An 80% performance accuracy was used as a learning threshold similar as in the training session.
(B) Reward gain by delay start. The reward gain was computed as the difference in the reward score between the delay and immediate start trials. This analysis showed that the reward score was significantly high in Condition 2 than in Condition 1 in the first half of the experiment.
(C) Percent reward score by trial type: error trials; suboptimal goal reach trials and optimal goal reach trials. There was a larger number of error trials in Condition 1 than in Condition 2 and a larger number of optimal goal reach trials in Condition 2 than in Condition 1.
(D) Cumulative reward score.
The legend for (C) and (D) are the same as in (A).

In a subsequent analysis, we classified the trials in three types based on the success in goal reaching: a) error goal reach - subjects failed to reach the goal; b) suboptimal goal reach - goal was reached by performing excessive finger movements, and c) optimal goal reach - goal was reached by performing an optimal action sequence. We checked the trial-by-trial time course and the frequency of occurrence of one of these trial types. Figure 5A shows the percentage of error goal reach trials and shows a fast decrease in error trials in Condition 2 in delay response start trials and a slow decrease in error trials in Condition 1. On the other hand, there was a slow and comparable decrease in error trials under the immediate response start in both Condition 1 and Condition 2. Figure 5B shows the percent of suboptimal goal reach trials. In Figure 5C we show the percent of optimal goal reach. In this graph, it can be seen that there was a fast increase in the number of optimal trials in Condition 2 under the delay response start, whereas the percent of optimal goal reach trials in Condition 1 was very low from the first trial and reach the same performance level of Condition 2 only in the 7th trial. Figure 5D shows the overall percent of the different trial types for each task condition. Subjects had more

error goal reach trials in Condition 1 than in Condition 2. On the contrary, Condition 2 had a larger number of optimal goal reach trials than in Condition 1. Figure 3E shows the cumulative reward score.

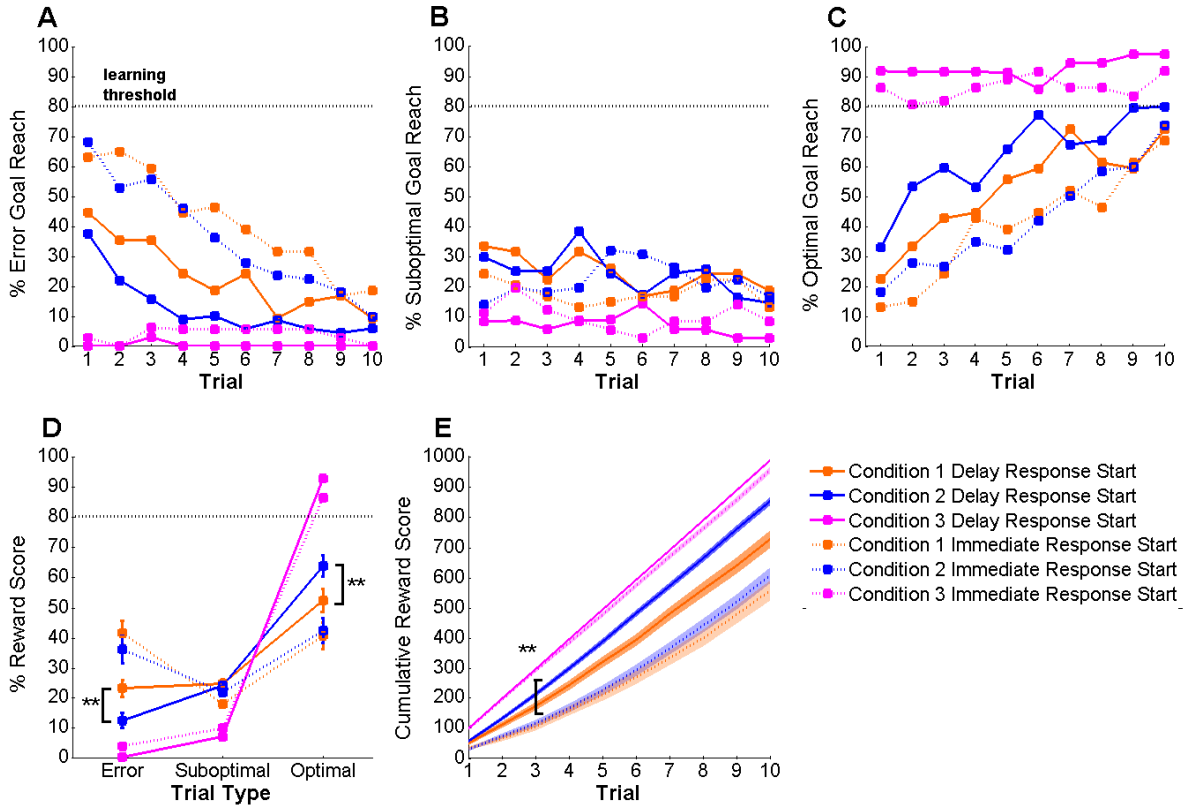

Figure 5. Analysis of performance by classification of trial types.
(A) Percent of error goal reach trials. In these trials subjects failed to reach the goal position.
(B) Percent of suboptimal goal reach trials. These are the trials in which subjects reached the goal position by performing excessive finger movements.
(C) Percent of optimal goal reach. In these trials subjects always reached the goal position by performing optimal (shortest paths) sequences of finger movements.
(D) Cumulative reward score.

We also analyzed the variability in the reward score by calculating the mean absolute deviation (MAD) for each subject and task condition, and entered this data into a group analysis using a three-way ANOVA with repeated measures. This analysis revealed a significant effect of test condition, $F(2,48)=130.44$, $p < 0.001$, start time, $F(1,48)=29.86$, $p < 0.001$, test block, $F(1,48)=34.93$, $p < 0.001$, and a two-way interaction for test condition and trial block, $F(2,48)=5.1$, $p < 0.01$. The post hoc analysis indicated that there was higher variability in the reward score of Condition 1 ($M = 36.66$, $SE = 1.34$) than in Condition 2 (30.39, SE 1.23) and Condition 3 ($M = 8.23$, $SE = 1.13$).

4.3.2.2. Reaction Time

The three-way ANOVA with repeated measures found a main effect of task condition, $F(2,3228)=72.85$, $p < 0.001$, start time, $F(1,3228)=2265.14$, $p < 0.001$, and trial block, $F(1,3228)=24.12$, $p < 0.001$, and for a two-way interaction between task condition and start time, $F(1,3228)=16.34$, $p < 0.001$, on the reaction time. No significant interaction interactions were found for test condition and test block, $F(1,3228)=2.8$, $p = 0.061$, start time and trial block, $F(1,3228)=0.4$, $p = 0.526$, and test condition x start time x trial block, $F(1,3228)=1.1$, $p = 0.332$. The post hoc analysis revealed no significant differences in the overall reaction time between Condition 1 ($M = 1.349$, $SE = 0.021$) and Condition 2 ($M = 1.383$, $SE = 0.018$), $p > 0.05$. The reaction time in Condition 3 was shorter than the reaction time in the other two conditions ($M = 1.014$, $SE = 0.026$), $p < 0.001$ (Figure 6A).

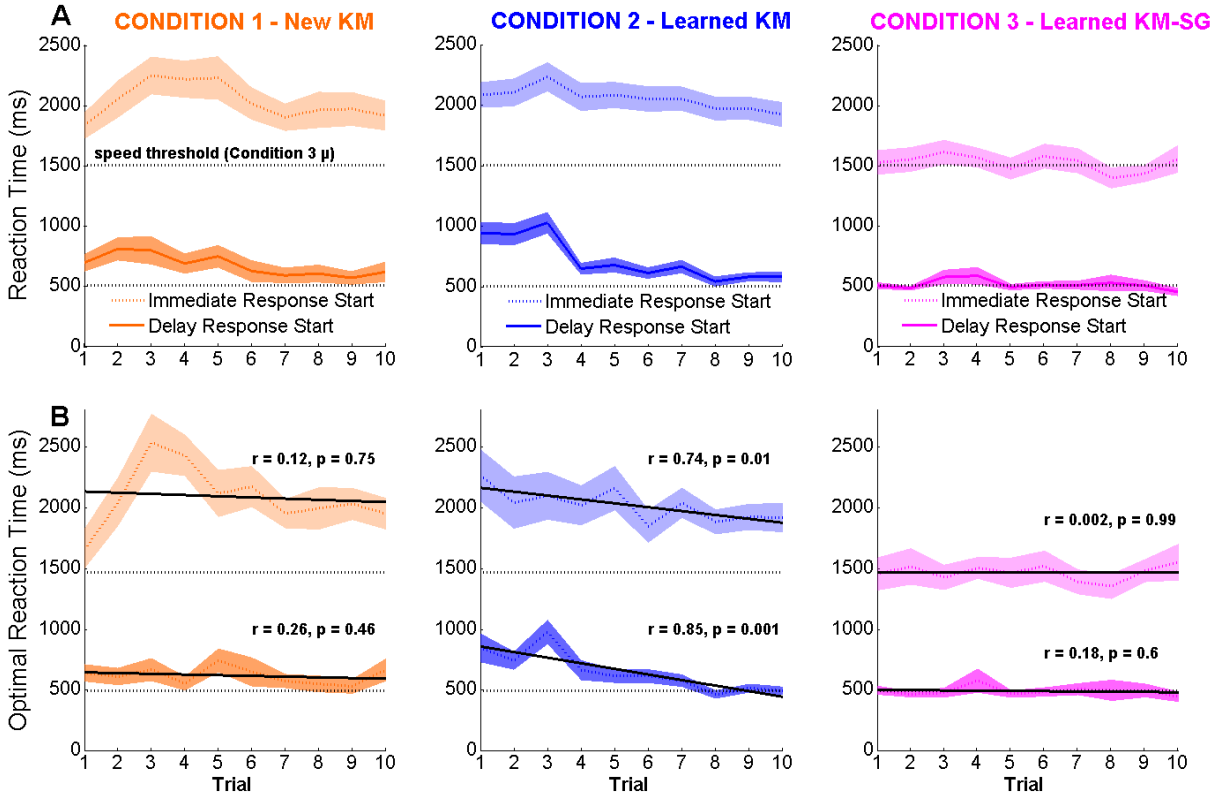

Figure 6. Reaction time performance.
(A) Reaction time including all trial types. The top and lower dotted lines represent the overall average reaction time of Condition 3.
(B) Reaction time including only the optimal goal reach trials. The top and lower dotted lines represent the overall average reaction time of the optimal goal reach trials in Condition 3.

A first-order polynomial function was used to perform a regression analysis and seek for any linear changes (increase of decrease) in the reaction time performance data. In this analysis only the reaction time of the optimal goal reach trials were used for linear regression. This analysis failed to identify any significant decrease in subjects' reaction time performance in Condition 1 in the immediate ($r = 0.12$, $p = 0.75$) and delay response start ($r = 0.26$, $p = 0.46$). However, we were able to find a significant decrease in the

reaction in Condition 2 under both immediate ($r = 0.74$, $p < 0.01$) and delay response starts ($r = 0.85$, $p < 0.001$, Figure 6B). No significant linear correlation was found in Condition 3 under the immediate ($r = 0.002$, $p = 0.99$) and delay response starts ($r = 0.18$, $p = 0.6$).

4.3.2.3. Execution Time

Figure 7A show the average execution time with all trial types included and figure 7B shows the average execution of the optimal goal reach trials. The three-way ANOVA with repeated measures found a main effect of task condition, $F(2,3221)=265.24$ $p < 0.001$, a nearly significant effect of start time, $F(1,3221)=4.69$, $p < 0.05$, and trial block, $F(1,3221)=203.65$, $p < 0.001$, and for a two-way interaction effect between task condition and trial block, $F(1,3221)=13.23$, $p < 0.001$. Condition 1 had an overall longer execution time ($M = 3.137$, $SE = 0.035$) than condition 2 ($M = 2.961$, SE 0.030), and condition 3 ($M = 1.912$, $SE = 0.043$).

A MAD analysis was used to seek for performance variability in the execution time using the same procedures as described in the analysis of reward score. The three-way ANOVA with repeated measure revealed a significant effect on the variability in the execution time by test condition, $F(2,48)=102.91$, $p < 0.001$, start time, $F(1,48)=35.61$, $p < 0.001$, trial block, $F(1,48)=10.62$, $p < 0.005$, a significant two-way interaction for test condition and start time, $F(2,3221)=11.29$, $p < 0.001$, and test condition and trial block, $F(2,48)=16.23$, $p < 0.001$. No other two- or three-way interaction effects were found. The post hoc analysis showed that the variability in execution time was much higher in

condition 1 (M = 1.07, SE = 0.02) than in condition 2 (M = 0.95, SE = 0.02) and condition 3 (M = 0.63, SE = 0.02).

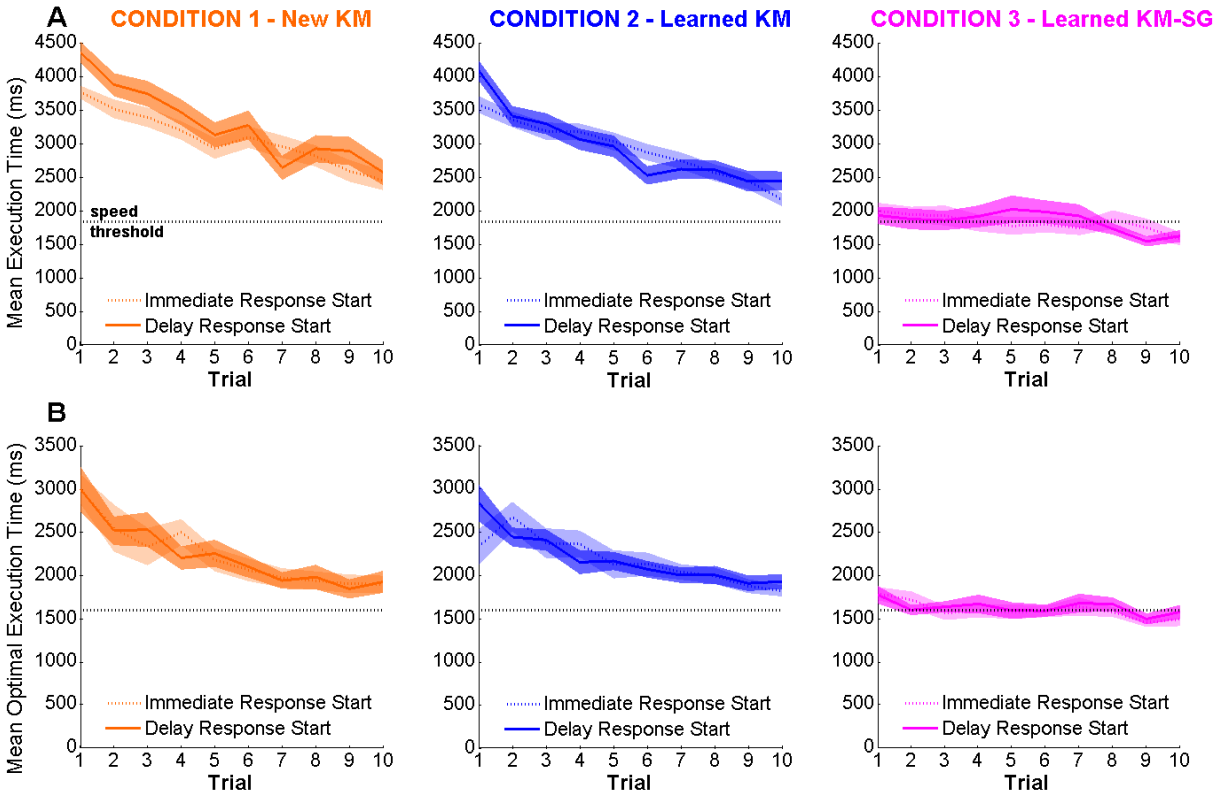

Figure 7. Execution time performance.
(A) Execution time including all trial types. The speed threshold (black dotted line) is the overall average reaction time of the delay response start trials in Condition 3.
(B) Execution time including only the optimal goal reach trials. The speed threshold (black dotted line) is the overall average reaction time of the optimal goal reach delay response start trials in Condition 3.

4.3.2.4. Number of Key-Presses to Reach a Goal

We carried a three-way ANOVA with repeated measures on the number of moves that subjects performed to reach a goal position in each task condition (Figure 8). We found a

main effect of task condition, $F(2,3220) = 87.11$ $p < 0.001$, start time, $F(1,3220) = 9.42$, $p < 0.005$, and trial block, $F(1,3220)=5.73$, $p < 0.05$, and for a two-way interaction effect between task condition and start time, $F(1,3220) = 3.64$, $p < 0.05$, and for start time and trial block, $F(1,3220) = 4.95$, $p < 0.05$. No other two-way of three-way interaction effects were found. The post hoc analysis indicated that the average number of steps in condition 1 ($M = 7.503$, $SE = 0.068$) was larger than in condition 2 ($M = 7.2$, $SE = 0.059$) and in condition 3 ($M = 6.1$, $SE = 0.083$).

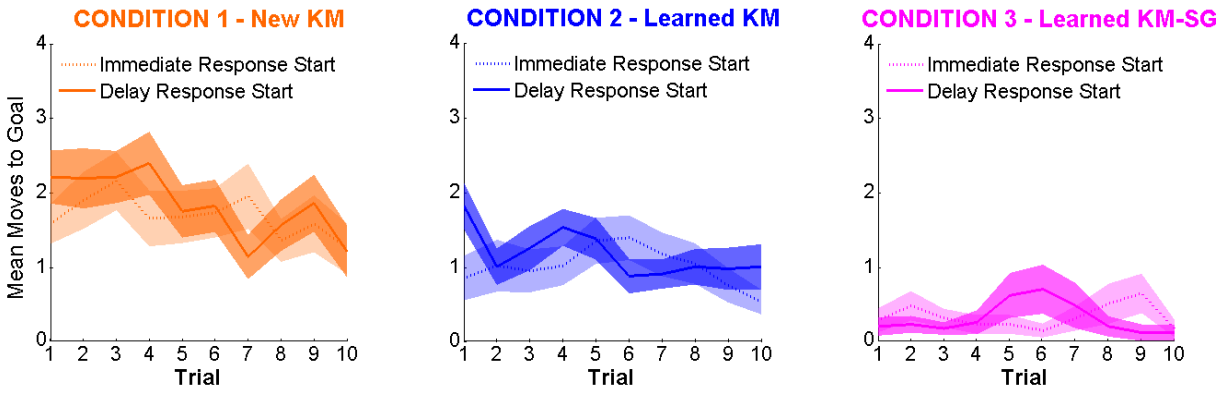

Figure 8. Number of moves. The number of moves was analyzed as the decrease in the number of extra finger button presses until performance reached the optimal number of moves to goal. The y-axis 0 value indicates a performance with no extra finger movements.

A MAD analysis was performed using the same procedures as described above. The three-way ANOVA with repeated measure revealed a significant effect of test condition, $F(2,48)=171.9$, $p < 0.001$, trial block, $F(1,48)=4.65$, $p < 0.05$, and a two-way interaction for test condition and trial block, $F(2,48)=8.71$, $p < 0.01$. No other two- or three-way interaction effects were found. The post hoc analysis showed that the variability in the

number of moves was higher in condition 1 (M = 2.01, SE = 0.06) than in condition 2 (M = 1.75, SE = 0.05) and condition 3 (M = 0.55, SE = 0.03).

4.4. fMRI Results

4.4.1. Delay-period Brain Activity in Condition 1 (New KM) and Condition 2 (Learned KM) vs. Condition 3 (Learned KM-SG sets)

Our main interest was in identifying how the different task conditions differentially modulated the BOLD fMRI signals from subjects during the delay period (4~6 seconds) of a trial, and to find out whether distinct neural networks were engaged in planning, selection and generation of motor responses. We evaluated subjects' elicited delay period BOLD signal from the entire test session, collapsed across the two scan runs and separately for each task condition. A subtraction analysis was performed using the delay period BOLD signal of Condition 3 as the control task to identify the areas that were significantly more active in Condition 1 (Condition 1 – Condition 3) and in Condition 2 (Condition 2 – Condition 3).

This analysis revealed a similar pattern of activation in brain areas in Condition 1 (Figure 9A) and Condition 2 (Figure 9B). Subtle differences, however, could be observed in a larger extension and spread of those activations in Condition 2 than in Condition 1 in

areas such as the dorsolateral prefrontal cortex bilaterally, ventral premotor cortex, lateral and dorsal premotor cortex, the basal ganglia, and left posterior cerebellum (see Table 1 for a complete list of activated brain regions). Significant differential activity was observed in the left anterior cerebellum in Condition 1, and the left posterior cerebellum in Condition 2.

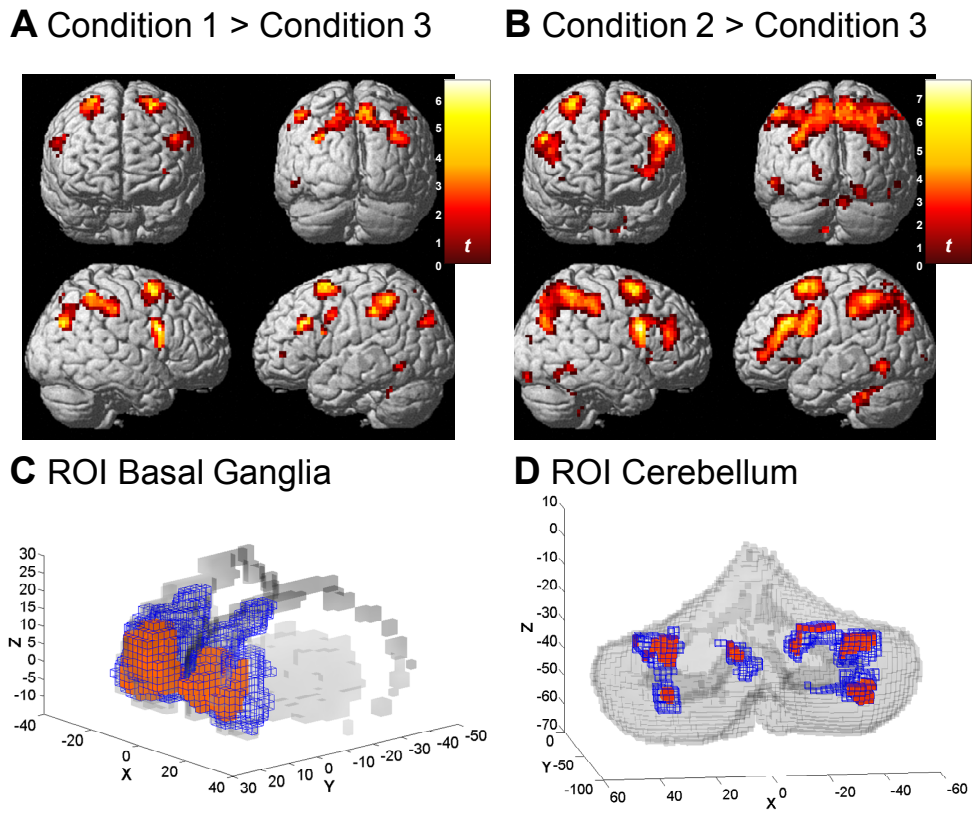

Figure 9. Delay-Period Activity in Condition 1 and Condition 2 After Subtraction of Condition 3

(A) Delay-period activity in Condition 1 (Condition 1 > Condition 3).

(B) Delay-period activity in Condition 2 (Condition 2 > Condition 3).

(C) ROI analysis using a mask image of the basal ganglia (caudate and putamen) extracted the activated voxels during the delay period for Condition 1 and Condition 2, as shown in (A) and (B), respectively. The basal ganglia activation in Condition 2 (blue color) extended from ventral to dorsomedial regions , whereas the activation in Condition 1 (orange color) was restricted to the ventromedial regional.

(D) ROI analysis, as described in (C), using a mask image of the cerebellum (restricted to areas Crus 1 and Crus 2). The extent of activation in the lateral hemispheres of the cerebellum was larger in Condition 2 than in Condition 1.

A ROI analysis on the basal ganglia showed a pattern of activation comparable to that observed in the cerebral cortex, the peak of activation in Condition 2 (Figure 9C) was on the left hemisphere (x = -12, y = 11, z = 1) and the activation extended in a rostro-caudal direction, from the more ventral to mid-dorsolateral caudate nucleus and putamen (voxel cluster size of whole ROI = 443), and the peak activation in Condition 1 (Figure 9D) was on the right hemisphere (x = 18, y = 11, z = 1) and located more ventro-medially including the caudate head and anterior putamen (voxel cluster size of whole ROI = 167).

4.4.2. Delay-Period Brain Activity for Condition 1 (New KM) vs. Condition 2 (Learned KM)

In the previous analysis we showed the presence of subtle differences in the regional patterns of brain activations for condition 1 and condition 2 during the delay period, which might be an indication of the existence of possible functional specificity operating for each task condition. Therefore, we conducted a subtraction analysis (condition 1 – condition 2; condition 2 – condition 1) to seek for areas whose activation were exclusively restricted to each of these two task conditions.

We found that condition 1 (Figure 10A) and condition 2 (Figure 10B) differentially relied on two distinct patterns of functional brain activity. Condition 1 primarily

activated the ventro-medial orbital frontal cortex, the somatosensory cortex, inferior parietal lobule, occipital cortex, and the posterior cerebellum, whereas areas activated in Condition 2 included the left dorsolateral prefrontal cortex, ventral premotor cortex bilaterally, right inferior temporal gyrus, left posterior cerebellum and a mid-dorsal region in the basal ganglia (caudate body). A ROI analysis identified two distinct clusters in the basal ganglia (Figure 10C), one in a ventro-medial region (peak: x = -12, y = 20, z = -5) in Condition 1, and the other in a mid-dorsal region (caudate body, peak: x = 9, y = 2, z = 16) in Condition 2.

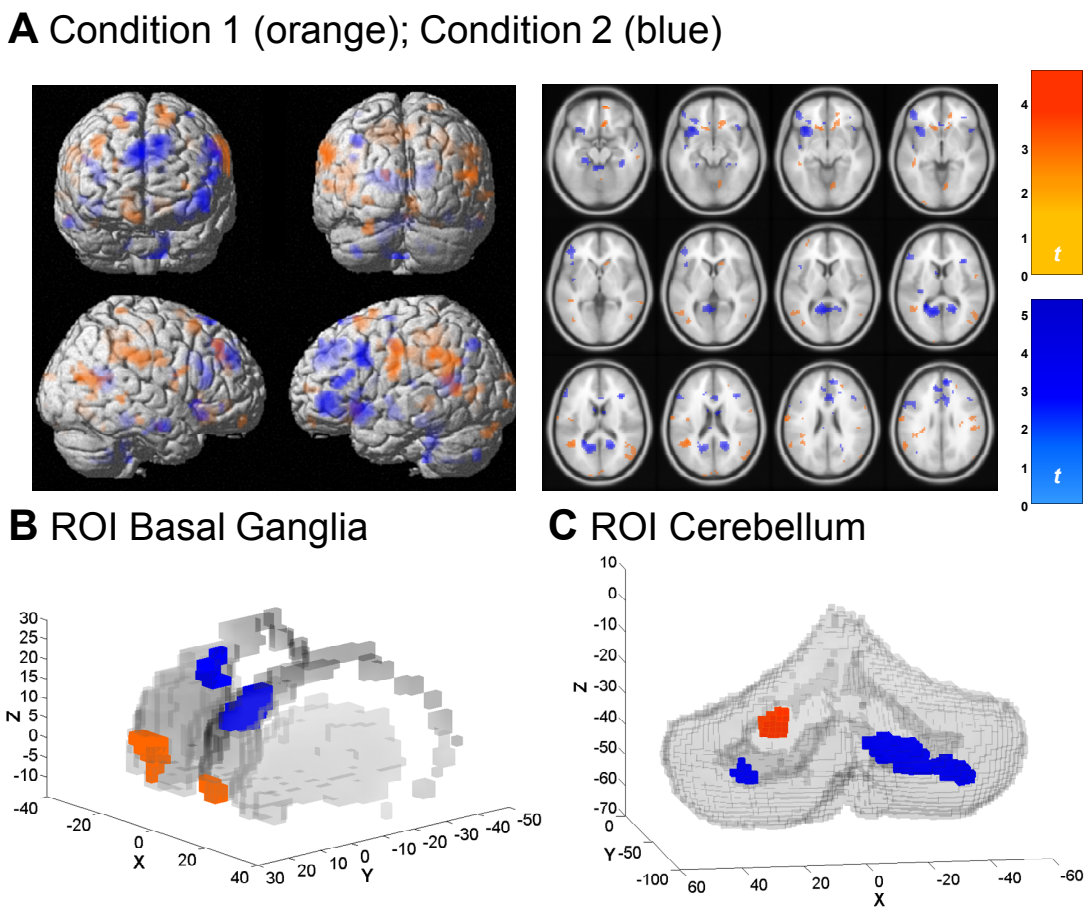

Figure 10. Delay-Specific in Condition 1 and Condition 2
(A) Delay-period activation after the subtraction Condition 1 > Condition 2 revealed that the areas activated in Condition 1 included mainly the medial orbital frontal cortex, the dorsal premotor cortex, the somatosensory and inferior parietal cortex. The areas

activated in Condition 2 after the subtraction of Condition 1 included the dorsolateral prefrontal cortex, ventrolateral premotor cortex, pre-supplementary motor area and superior parietal cortex.
(B) The mutual subtraction of the delay-period activity between Condition 1 and Condition 2 revealed separate sites of activation in the basal ganglia. Condition 1 activated mainly the anterior pole and ventromedial region, whereas the dorsomedial caudate nucleus was activated in Condition 2.
(C) The mutual subtraction also showed the activation of distinct regions in the cerebellum. The left lateral region of Crus 1 was activated in Condition 1 while the right Crus 2 with activation extending from the vermis to the lateral part was activated in Condition 2.

4.4.3. Delay-Period Brain Activity in Condition 3

We were also interested in the neural mechanisms associated with the automatic, habit-like performance of task condition 3, and for that purpose, a further analysis was conducted to investigate the fMRI signal during the delay period using a series of simple subtractions (Condition 3 – Condition 1, Condition 3 – Condition 2) and two-way subtractions (Condition 3 – [Condition 1 + Condition 2]). This analysis allowed us to confirm the consistency of the results previously reported (see Figure 4) when condition 3 delay period activity was used as a baseline control and to see whether other brain regions also were specifically activated when subjects performed the condition 3 (Figure 11).

A B

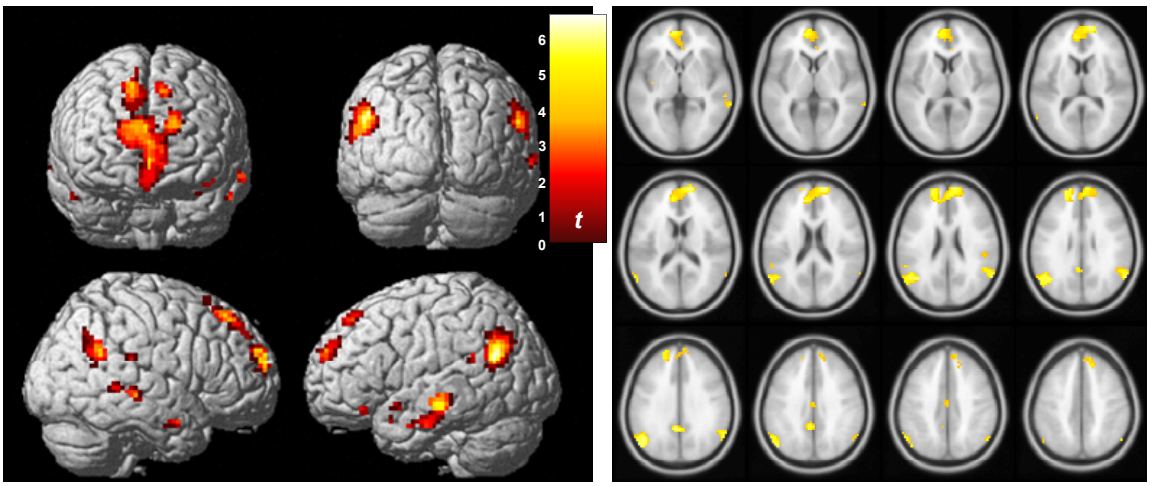

Figure 11. Brain activity related to retrieval of motor memories
Activation in Condition 3 during the delay period after a two-way subtraction (Condition 3 - [Condition 1 + Condition 2]. There was activation in the

The analysis revealed that there was a common pattern of brain activation, across the three types of subtractions implemented, in the superior medial frontal gyrus, middle temporal gyrus, angular gyrus, anterior cingulate, and putamen (Figure 11). In the results presented in Figure 9, the delay period activity was used as a control task. Interestingly, when the delay-period activity of Condition 3 is compared with the baseline, the brain areas that are significantly activated include the supplementary motor area, lateral premotor cortex, primary motor cortex, somatosensory cortex, precuneus, anterior and posterior motor cerebellum and posterior putamen (Figure 12).

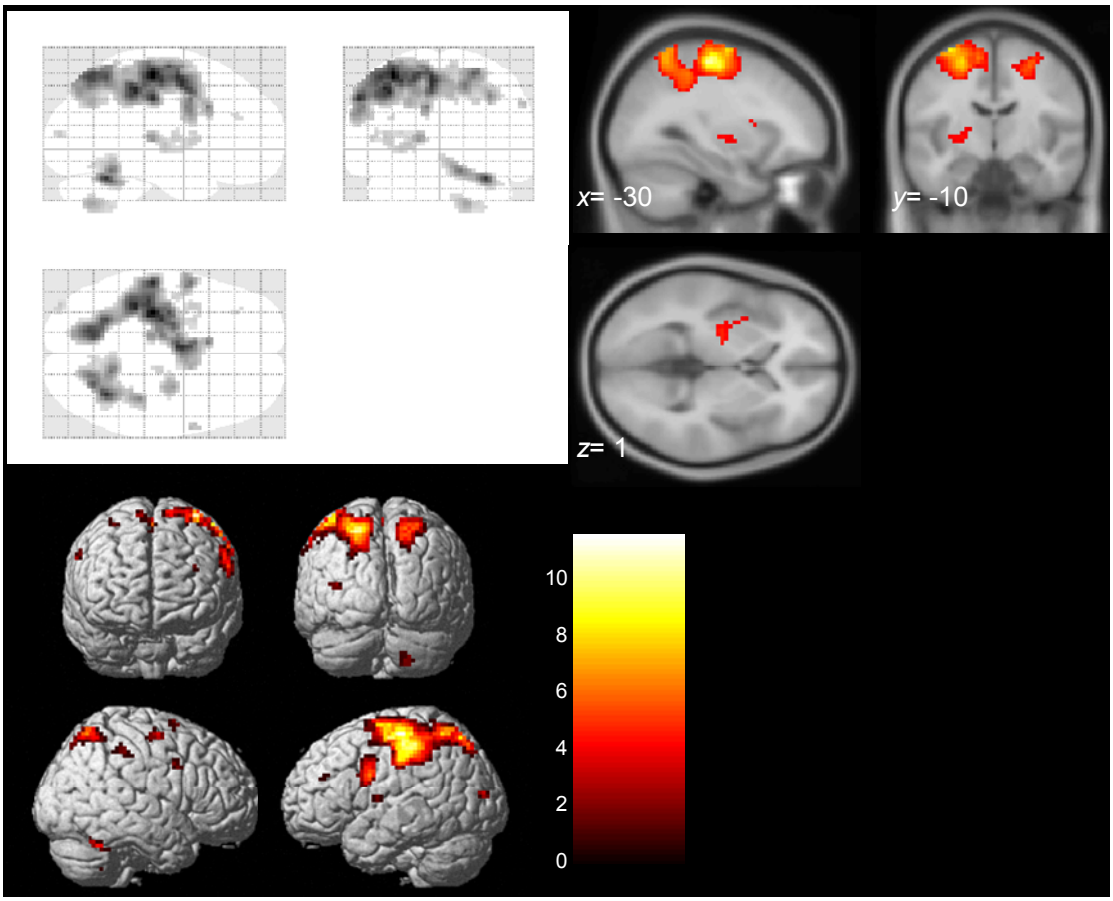

Figure 12. Delay-period activity in Condition 3 vs. baseline.
The subtraction of delay-period activity in Condition 3 > baseline revealed a pattern of brain activation that included mainly sensory-motor areas involved in storage and implementation of habitual, automatic motor behaviors: supplementary motor area, lateral premotor cortex, primary motor cortex, somatosensory cortex, precuneus, posterior putamen (dorsolateral striatum) and anterior and posterior motor cerebellum.

We also conducted the same type of two-way subtractions for conditions 1 and 2, and found a clear separation of areas active for each condition. In condition 1 the activation was restricted to the premotor cortex (BA 6), right inferior operculum, inferior parietal gyrus (BA 40), precuneus (BA 7) bilaterally, and left ventral striatum, while in condition 2 the active areas were the dorsolateral prefrontal cortex (BA 46), left premotor cortex

(BA 6, BA 9), left supplementary motor area, posterior cerebellum, caudate body, internal globus pallidus, and middle occipital gyrus.

A ROI analysis on the basal ganglia using this two-way subtraction method for the three task conditions revealed a gradient of activation that seemed dependent on the level of experience, where a more ventral striatum was activated in condition 1, a mid-dorsal region (centered on the caudate body) was activated in condition 2, and in condition 3 a dorsolateral region (putamen) was active. The single subject contrast images from the two-way subtraction were entered in a conjunction analysis, so that we could fully demonstrate that the three task conditions elicited activation in distinct networks. The results of the conjunction analysis showed that no regions were commonly activated among the three task conditions.

4.5. Discussion

In this study, we had human subjects perform a grid-sailing task to learn key mappings and action sequences to move a cursor from its start position to a target goal position. After one day training and one night sleep interval, subjects were tested under three task conditions that, depending on their familiarity with the key mappings, mimicked different levels of experience in action selection.

The analysis of subjects' behavior in the training session showed significant learning-related improvements with a gradual increase in the reward score and decreases in the number of moves performed to reach a goal position, in the reaction time and execution time. Although significant differences were found among training groups for specific behavioral measures, however, such differences completely disappeared with a one-night sleep interval as indicated by the analysis of the reward score during the pre-fMRI short practice. These results confirmed that subjects successfully learned the task, and the formation and consolidation of automatic, habit-like action sequences (Hikosaka et al, 1995; Adams, 1971).

Behavior analysis in the test session revealed distinct performance profiles for each task condition. Significant learning-related changes were found in conditions 1 and 2 but not in condition 3. The highest variability in all performance measures was found in condition 1, which also was associated with the lowest reward score, longest execution time, and larger number of moves. These results are in accordance with the motor learning literature suggesting that novel action learning is characterized by the use of exploratory behaviors to capture the dynamics of the interaction between actions and the environment, consequently leading to an initial low performance and high variability in behavior (Gabriele, 1981; Hall, 1990; Hyman, 1993).

The results also provide evidence to our hypothesis that, in the absence of an internal model to plan and predict an action outcome, humans utilize a model-free action value-based strategy for action selection. The long execution time can be explained by the large number of moves, by a longer time to generate the next action, and also by the necessity

to receive visual feedback of cursor motion to establish associations between finger movement and cursor motion direction. Interestingly, the increased reaction time in the immediate start in the first trial block might be an indication that the internal model of the new key mapping was under construction.

Behavioral performance in task condition 2 was characterized by a boosted learning with a fast increase in the reward score, especially when the response started after a delay, which indicates a beneficial effect of this extra time on the planning of future action sequences. We also found that the reward gain was greatest in condition 2, which was correlated with the highest probability of performance of optimal action sequences even from the very first trial of the experiment. In addition, the overall performance in condition 2 was significantly more consistent and less variable than in condition 1.

The beneficial effect of a delay period on motor performance is well known, for example, in a classical study, Sheridan (1989) tested human subjects performing stereotypical sequences of arm-hand movements under several start time and advance information conditions, and found that speed and quality of performance improved when the response started after several seconds, or when information indicating the motor sequence to be performed was given in advance. The main difference between Sheridan's paradigm and our condition 2, however, is that in our experiment subjects started off the task knowing only the key mapping rules and cursors, but not the correct action sequences, which had to be learned on the fly during the fMRI scanning. Although the actual action responses were unknown, our results showed that subjects could successfully utilize the available internal model of the pre-learned key mappings to plan future action sequences. These

results are in agreement with our hypothesis that given the availability of time for computation and the existence of an internal model, human subjects can implement a model-based forward model strategy by mental simulation to predict and evaluate state transitions and hypothetical actions, and to select those actions that are most likely to lead to successful outcomes.

In task condition 3, subjects performed action sequences, after the passage of a one-night sleep interval, which had been well-learned on the previous day in the training session. We found that performance in condition 3 had the highest reward score, which was unaffected by the response start time, and a fast reaction time and execution time. There were no significant learning- or performance-related improvements in this condition. Although the action sequences have been practiced for just one day and retested on the following day, it might be difficult to assume they already were in a habitual mode. However, given the high reward score performance, the fast execution and low variability, the results provide support to our hypothesis that an automatic sequential motor memory process in the form of a learned policy of model-free algorithm guided the reproduction of the learned action sequences.

The behavioral results suggest that in condition 1, characterized by exploratory behavior, subjects may have utilized an action value-based model-free strategy for action selection; in condition 2 the boosted learning and large reward gain in delay start trials suggest the use of a model-based forward model strategy for prediction, evaluation and selection of actions, and the high performance in condition 3 indicates the use of a motor memory policy of model-free algorithm to reproduce learned action sequences.

The fMRI imaging data also showed that each of these task conditions associated action selection strategies had a counter part in the brain mechanisms implementing them. Condition 1 activated a network linking mainly the orbitofrontal cortex and anterior ventral striatum, whereas the dorsolateral prefrontal cortex, ventral premotor cortex, mid-dorsal striatum and lateral cerebellum were activated in condition 2, and the superior medial, middle temporal gyrus and dorsolateral striatum were activated in condition 3.

4.5.1 Related Sequence Learning Studies

Our study was the first to use a behavioral task that tried to explicitly separate the cognitive processes, expected to be predominantly engaged during the delay period (4~6s), from the movement execution period after the start go signal. We cannot rule out, however, that abstract motor information was used during the delay period, since the internal model, represented by the key-mapping rule, was associated with the movement of specific right hand fingers as well as the motion direction of a cursor. Other studies, on the other hand, have investigated the process of sequence learning and related brain structures using fMRI and PET by experiments in which subjects had to learn online by movement execution and trial-and-error the correct finger sequence.

The first behavioral paradigm for investigating sequence learning was based on the serial reaction time (SRT) task. In this paradigm subjects are informed that one of four parallel horizontally arranged stimuli will turn on and they have to press the corresponding button (such as index, middle, ring and little fingers). The lights sequentially turn on for

a period of about 60s, but unknown to subjects, the lights follow a predetermined sequence of eight moves. Some of the subjects eventually learn the hidden sequence and decrease the reaction time for button pressing. Other subjects never find or cannot report the hidden sequence at the end of the experiment, although their performances measured by the reaction time also improves, that is, button press becomes faster but still significantly slower than the subjects who explicit learn the sequence.

Interestingly, the areas activated in group of explicit learners during the initial learning stage is similar to those activated in our experiments in Condition 1 and Condition 2, including the ventromedial prefrontal cortex, dorsolateral prefrontal cortex, supplementary motor area and lateral cerebellar lobes. The areas activated in the group of implicit learners include mainly sensory-motor related areas, the supplementary motor area, primary motor cortex and posterior cingulate and parietal cortex. This pattern of activity is in line with the results found in our Condition 3 (learned KM-SG sets) which had predominant activation in the supplementary motor area, primary motor cortex, precuneus, dorsolateral striatum (posterior putamen) and anterior and posterior motor cerebellum.

These results suggest that explicit learning requires the use of brain areas responsible for the control of goal-directed motor learning which subsequently might lead to better improvements in motor performance, whereas the areas activated in implicit learning are involved mainly in automatic or reactive motor processes that require no learning although motor improvements may still occur. The use of SRT tasks, however, cannot clearly separate planning and motor processes since there is no explicit delay before

movement execution and subjects do not need to use abstract rules of other cognitive processes for sequence generation. Another difference with our study is that a SRT task does not require learning by trial-and-error since the appropriate motor response are determined in advance by the experimenter.

The works by Hikosaka and Passigham are the ones that share one important feature with our study, the learning by trial-and-error using a reinforcement learning paradigm. In Hikosaka et al (1999) 2x5 task, monkeys and humans have to push in the correct sequence one of 10 buttons, arranged in a button pad of 4x4 LEDs. The buttons turn on to indicate the go signal and if a button is pressed incorrectly, all buttons turn off and the sequence has to start from the beginning with no delay. In this task the visual signal works as reinforcement feedback. The works by Passingham follow the same learning paradigm but auditory feedback is used instead of a visual signals.

In such reinforcement learning paradigms, subjects have to actively search the correct sequence of finger movements, therefore, exploration and memory of the correct movements must be kept in mind, as well as corrections to the wrong movements and anticipation of next movements. Subjects are put in a real problem-solving learning task. These experiments have found a group of brain areas is predominantly activated in the initial cognitive goal-directed learning period including the dorsolateral prefrontal cortex, pre-supplementary motor area, posterior parietal cortex, anterior basal ganglia and lateral cerebellum. The activity in these areas decrease with advanced practice and another group of areas become predominantly active when performance reaches an automatic/habitual level: supplementary motor area, lateral premotor cortex, precuneus,

anterior cerebellum and posterior putamen. These were the first studies to explicitly show that different neural circuits take over behavioral control in different learning stages. These results were supported by Hikosaka's lesion studies with monkeys (see Chapter 2 for details). A difference Hikosaka's an d Passingham's studies with the task reported in this thesis is absence of an explicit delay period for advance response preparation, and the brain activity is averaged over periods that include many actual finger movements.

Mushiake et al. (2001) have performed a fMRI experiment with humans using a path planning task in which subjects had to use specific rules, indicated by the cursor color, to plan a sequence of finger movements to move a cursor from its start position to a target goal. The task is very similar to the one presented here and starts the presentation of the start-goal positions and the cursor (with a color indicating the appropriate rule) followed by a delay period of 6 seconds. Therefore, by using this task Mushiake and colleagues were able to explicitly identify the brain areas activated for planning of movements ahead. However, subjects were highly trained to perform this task which also required the planning of quite short movement sequences, only 3 steps to goal. Consequently, very little activity was found in brain areas related to high-order cognitive functions and sequential planning, such as the dorsolateral prefrontal cortex. Instead, what these researchers found was a pattern of brain activity that resembled the one identified in our task Condition 3. The areas activated during the delay period were the supplementary motor area, superior parietal and precuneus and anterior cerebellum. Therefore, in this case, the task elicited activity mainly in sensory-motor areas even to the planning of new but short easy movement sequences. In our task we could clearly test subjects' capacity

of learning, planning and generate reactive behaviors by testing them under task conditions that required different levels of experience. In Mushiake's study, the task might have been too easy to force subjects to use more complex cognitive processes for planning a sequence of movements.

The present study was not concerned with the mechanisms involved in the temporal and sequential organization of motor behaviors. Tanji and his colleagues (Ninokura, Mushiake and Tanji, 2003; Mushiake et al, 2006; Mushiake et al, 2009; Saito et al, 2005; Shima et al, 2007, Tanji and Shima, 1994; Tanji, Shima and Mushiake, 1996; Tanji and Hoshi, 2001; Tanji, 2001) using similar maze-navigation or three-turn wrist movement tasks have been able to identify some basic neural mechanisms for planning and sequencing of hand movements. In the navigation task requiring path planning for different start-goal positions and in trials where certain possible maze paths were blocked with obstacles, activity in the dorsolateral prefrontal cortex correlated with the planning and representation of immediate and final behavioral goals (Ninokura) as well as multiple steps of future events (Mushiake et al, 2006). On the other hand, neurons in the supplementary motor area were found to code for single hand movements or specific transitions between movement sequences (Tanji, 2001; Tanji and Shima, 1994; Shima et al, 2007). These results are in agreement with the findings of other studies with monkeys (Fujii and Graybiel, 2003) which identified that the activity some neurons in the prefrontal cortex code sequence boundaries, that is, the start and end of a motor sequence. Similar results were obtained by Xin and Costa (2010) who found that nigrostriatal neurons code for the start/stop of motor sequences. Jubault, Ody and Koechlin (2007) and Koechlin and Jubault (2006) also found, using fMRI with humans, that the activity

in the ventrolateral prefrontal cortex (Broca's area) codes the start and end of sequences, whereas the activity in the posterior parietal cortex codes the transition within elements of a sequence.

4.5.2. Reinforcement Learning Paradigms for Learning, Planning and Habitual Behaviors

We have designed our grid-sailing task and analyzed the behavioral and brain imaging data in light of Reinforcement Learning (RL) theory and its possible biological implementations in the human brain (Sutton and Barto, 1998; Doya, 2007; Daw, Niv and Dayan, 2005; Montague et al, 1996; Ito and Doya, 2011). In this section I will discuss the results presented in this thesis with other studies which used RL or other computational methods to model subjects' behaviors and the computational roles played by distinct brain regions.

In one of such works Yoshida and Ishii (2007) have conducted an elegant experiment and used a sophisticated computational model to analyze subjects' sequential decisions. These authors developed a navigation task for which subjects were shown only once the target goal in a 7 x 7 grid environment. After the disappearance of the grid, subjects could only navigate towards the goal if they correctly estimated the current position based on 3D scenes indicating their possible locations in the grid maze. A Hidden Markov Model (HMM) was created to estimate by using Bayesian inference subjects' internal cognitive processes based on their choice history. The task itself is a difficult problem to solve but subjects eventually became proficient at finding the optimal path to

goal. In addition, a unique pair of start (unknown) and goal positions was tested for every trial, therefore, the task did not require the learning or acquisition of stereotypical behaviors such as in our task Condition 3 (learned KM-SG set). The authors found that dorsolateral prefrontal cortex, the anterior cingulate cortex, the posterior parietal cortex and the anterior basal ganglia correlated with the planning of the forthcoming decision. The regression analysis using parameters of the HMM identified the medial prefrontal cortex as responsible for estimation of current estate within the maze.

These results provide support to our findings in task Condition 2 (learned KM new SG) which showed that similar brain areas were activated when subjects were had the chance to use the learned KM during the delay period to plan ahead a sequence of movements to reach a goal position. In that regard, in Wako-Ishii task subjects could use the 3D scene diagram to estimate their current state and plan the next action in the same way that our subjects could use the learned KM to plan a sequence of finger movement given their initial state at the grid. In their experiment, subjects became faster at finding the goal position after repeated experiences even for new start-goal positions, which might indicate that subjects could successfully implement some sort of model-based forward planning, such as in our task Condition 2, by using the 3D scene diagrams. A difference in their task compared with ours is the absence of a delay period, except for a short period of about 2s that their subjects had to make a the next decision. In addition, similar to other sequence learning paradigms, the brain activity was averaged over a long period which included several cognitive and motor processes, and therefore, brain areas other than those found in our results also showed up in their findings.

Glascher et al. (2010) have devised a tree-search two-choice probabilistic task in order to understand what possible reinforcement learning computations might be implemented by distinct cortical and subcortical regions for decision making. These authors speculated that the brain might utilize two types of prediction errors, one for state prediction and the other for reward prediction. State prediction error concerns the prediction of which state is expected to be achieved given the current state, whereas reward prediction error is the difference between the actual and expected reward for a given state. Glascher and his colleagues found that the activity in the dorsolateral prefrontal cortex best correlated with a state prediction error function, and the ventral striatum activity correlated with computing the reward prediction error. In our study we have not yet analyzed what computational functions each of the elements in the prefrontal-basal ganglia and cerebellar networks play in the process of learning, planning and generation of automatic behaviors. However, the results found by Glascher et al (2010) provide substantial support for our hypothesis that the dorsolateral prefrontal cortex might implement some kind of state prediction and state transition computations. It is not clear, however, how this region computes state prediction error, whether locally or in coordination with other brain areas, including the basal ganglia.

We have to far discussed the possible role of the dorsolateral prefrontal cortex in planning of future events, such as reaching a final target goal or prediction of the next state given the current sensory state and available actions. The literature linking different parts of the prefrontal cortex in higher-order cognitive processes is now well-established from the view point of cognition (see Chapter 2). However, the computational roles

played by these regions for planning, as well as for learning or the control of habitual actions is still an open issue.

Chapter 5

Conclusions

We could demonstrate that humans use multiple action selection strategies depending on their experiences, existence of an appropriate model for the task and available time for thinking. The fMRI results also showed that the use of distinct strategies for action selection relied predominantly on the activity of specific neural networks including the prefrontal cortex, basal ganglia and cerebellum. With the use of the concept of transfer of learning and our assumption that the cognitive processes engaged in transfer are similar to those happening in an intermediate stage of learning, we were able for the first time to break apart the different stages of learning, and testing these stages separately under different task conditions.

We also demonstrated the common principles involved in action learning and decision making, where the initial stage of motor learning corresponds to learning to make decisions from scratch, the intermediate stage of learning corresponds to the use of thoughtful deliberative model-based strategies for planning future behaviors, including the estimation of state transitions and covert multistep behaviors, and the advanced stage of motor learning correspond to the selection of well-known stimulus-response associations.

Evolution has shaped the human brain with multiple action selection strategies so that optimal behavior can be achieved in multitasking situations. In a situation where a model of the environment is present, it is not necessary to relearn the same model every time the same situation is encountered. Therefore, we bypass the initial stage of learning. The same applies for the actions that have been well-mapped to a given situation. It's enough to retrieve the motor memory and perform the action. The use of this kind of memory system of the brain frees the other remaining action selection systems for engagement if the situation requires the combination of habitual actions, and the generation of more cognitive controlled actions.

5.1. Summary of Contributions

The results of the two behavioral experiments and those of the fMRI study allowed us to make the following contributions:

I) Evidence on the use of model-based strategy for action selection.

The behavior results of the experiment described in Chapter 3 demonstrated that subjects used a model-based strategy for planning sequences of finger movements. The evidence confirming our hypothesis was provided with the large reward gain by the delay start in task condition 2. This means that subjects could make use of previously learned KM during the delay period plan and prepare an action sequence by mental simulation.

II) Action selection ability is experience-dependent.

In the second behavioral experiment, presented in Chapter 3, the task design was modified to make it methodologically and statistically valid. Subjects were tested in the main experiment under three task conditions, such as learning from scratch (Condition 1), action planning with the existence of prior knowledge (Condition 2) and automatic/habitual actions (Condition 3). We found distinct performance profiles under each task condition, a slow and variable exploratory performance in Condition 1, boosted performance and larger reward gain in Condition 2, and a fast and accurate performance in Condition 3. These results suggest that different action selection strategies were used for solving these multiple tasks.

III) A cascade parallel model for action selection and learning.

Base on the behavioral results, we proposed a model for action selection and learning using concepts from reinforcement learning theory. The cascade parallel model composed of three systems is able to explain why stage-wise transitions are observed in motor learning. Since the three systems operate in parallel, their engagement in learning is conditional to several task factors, such as task difficulty, experience with the task and existence of internal models or necessity of acquisition of the task dynamics.

IV) Multiple models for action selection.

Based on the results in (**II**) we proposed formal methods for action selection based on the theory of reinforcement learning, namely: a value-based method is used for exploration and acquisition of the task dynamics, a model-based method is used when the model of the task dynamics is available for planning, simulation and prediction of action outcomes

and state transitions, and a memory-based method is used for the quick selection of well-learned sensory-motor mappings such as habitual actions.

V) Shared principles for action selection in motor learning and decision making. The grid-sailing task was designed to provide an interface and test the problem of action selection in decision making and motor learning paradigms. We assumed that the different task conditions (Condition 1, Condition 2, and Condition 3) provided this interface by breaking apart the different motor learning stages (early, intermediate and late), which also mimicked different levels of experience in action selection and decision making (unknown/novel task, thoughtful purposive action planning, habitual/reactive).

VI) Time is essential for the realization of model-based strategies.

A drawback in the implementation of model-based strategy is that it is timing consuming, its success depends on the model accuracy and memory load. In actual daily life humans face many situations requiring an immediate response which sometimes turns out to be bad even though we knew the outcome could have been better if given more time. In the grid-sailing task we introduced a delay period for planning that preceded the go signal, and demonstrated that performance was boosted especially when subjects had previous experience with the task and could use that experience for successfully planning action sequences (Condition 2). These results have education and therapeutical applications. Because of the lack of experience, students many times need more time to process and get familiar with new information provided by the teachers. The same applies for patients going through cognitive therapies or physical rehabilitation.

VII) Multiple prediction models for action selection are present in prefrontal-basal ganglia and cerebellar networks.

The analysis of the fMRI data time locked to the brain activity during the delay period showed that distinct neural networks were active as subjects performed the different tasks. In Condition 1, when behavior was exploratory and variable, which mimicked the initial stage of motor learning, and used an action-valued based strategy for action selection, the medial orbitofrontal cortex and ventromedial striatum were preferentially active. On the other hand, the boosted learning by the use of a planning model-based strategy in Condition 2 predominantly elicited activity in the dorsolateral prefrontal cortex, dorsomedial striatum and right lateral cerebellum. Finally, the use of a memory-based method for the fast selection and implementation of habitual/automatic actions in Condition 3 was associated with activity in the supplementary motor area and posterior dorsolateral striatum. These results are supported by anatomical studies showing reciprocal connections between the medial orbitofrontal cortex and ventromedial striatum, the dorsolateral prefrontal cortex and dorsomedial striatum, and dorsolateral prefrontal cortex and lateral cerebellum, and the connections between the supplementary motor are and dorsolateral striatum. Thus, the results provide evidence that these three parallel neural networks are likely the neural substrates for implementing distinct reinforcement learning action selection strategies under different task requirements, such as for learning by exploration, by planning or simulation, and the selection of habitual motor memories.

5.2. Future Directions

The behavioral and fMRI results presented in this thesis were able to shed light several important topics for the problem of action selection in learning and decision making contexts. However, several behavioral, computational and neuroscience problems still need to be elucidated.

I) What are the processes involved in the acquisition of a model of the task dynamics?

It is always a difficult problem for humans and animals to implement a behavior policy in a novel, completely unknown task. The task, therefore, should be learned from scratch by exploration and trial-and-error if a reward signal is available to evaluate the overt motor behavior. However, it is not know what mental processes are involved in such exploratory knowledge-free learning situations, nor how and what information is picked up from the environment and assembled together to form a model of the task.

II) What are the internal mechanisms and brain architectures for planning and mental simulation?

In many learning or decision making situations humans can make use of previous experiences to better perform at the task at hand. The quality of the performance also depends on the model accuracy. For example, an experienced chess master is able to anticipate several moves ahead depending on the current board move made by his opponent. What are the mental strategies that allow humans to contemplate the future

and fictive situations, plan a sequence of responses and keep these responses in memory in the right order for later execution, simulate and anticipate the behavior of others, and evaluate the consequences of candidate actions? A remarkable feature of patients with schizophrenia, for instance, is the disturbance in their capacity to organize their thoughts in a meaningful manner leading hallucinations and delusional thoughts. Hence, understanding the normal mechanisms in mental simulation, its architecture and component parts that implement it is essential for developing alternative therapies for patients with neurological and developmental disorders.

III) Where in the brain are motor memories stored?

This is an old, but still highly controversial problem in human motor behavior. One possible reason is that experimentalists use very different tasks to assess motor memory formation and its counterpart brain mechanism. Another reason is the use of different methods for analysis of brain data. Finally, there is a debate of whether memories are stored in single brain regions or it is shared by distributed large-scale networks.

IV) Where in the brain are the internal models stored?

Similar to the problem of where in the brain motor memories are stored, there has been no formal test to find how and where internal models of task dynamics and environments are formed and stored. Internal models can be more complex than motor memories and made up of multiple pieces including perceptual, sensorial and somato-motor representations. Finding the precise location of where internal models are stored in the brain is a hard problem to be solved since these representations are highly distributed in the nervous system. Instead of local representation, it is more likely that a neural network

models might best explain where internal models are stored and how they are used for production of cognitive and motor outputs.

V) How does stage transition and switching of strategies occur?

The behavioral results and our proposed cascade model presented in this thesis suggest that the stage-wise transitions observed in human motor learning can be explained by a switching to an action selection strategy that is most adequate at given period of the learning process. humans use distinct action selection strategies. Our results also suggest the use of distinct selection methods in decision making. Nevertheless, it is not clear what factors influence how and when action selection switching takes place. In addition, as the multiple selection methods operate in parallel, it is not known which system preferentially generates the decision output.

VI) How to the multiple cortico-subcortical network loops communicate and share information?

We have shown that one of three distinct prefrontal-basal ganglia and cerebellar neural networks was predominantly active for a task condition. These results are supported by a neuroanatomical theory (Alexander et al, 1986) that suggests the brain is organized in parallel segregated network loops. This theory, however, fails to explain how these multiple parallel loops communicate and share information, which is essential to allow flexible switching to an action selection strategy to take over behavioral control.

REFERENCES

Abeele, S & Bock, O (2001). Mechanisms for sensorimotor adaptation to rotated visual input. Experimental Brain Research, 139: 248-253.

Ackerman, P. L. Determinants of individual differences during skill acquisition: cognitive abilities and information processing. Journal of Experimental Psychology: General, 117, 288-318, 1988.

Adams, J. A (1971). A closed-loop theory of motor learning. Journal of Motor Behavior, 3(2):111-49.

Adams, J. A., Marshall, P. H., Bray, N. (1971). Closed-loop theory and long-term retention. Journal of Experimental Psychology, 90(2), 242-250.

Adams, J. A. (1977). Feedback theory of how joint receptors regulate the timing and positioning of a limb. Psychological Review, 84(6), 504-523.

Adams, J. A. (1968). Response feedback and learning. Psychological Bulletin, 70(6), 486-504.

Adams, J. A., Bray, N. W. (1970). A closed-loop theory of paired-associate verbal learning. Psychological Review, 77, 385-405.

Adams, J. A. (1984). Learning of movement sequences. Psychological Bulletin, 96, 3-28.

Addis, D. R., Wong, A. T., Schacter, D. L. (2007). Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. Neuropsychologia, 45, 1363-1377.

Adolph, K, E., Vereijken, B., Denny, M. A. (1998). Learning to crawl. Child Development, 69(5), 1299-1312;

Akkal, D., Dum, R. P., Strick, P. L. (2007). Supplementary motor area and presupplementary motro area: targets of basal ganglia and cerebellar output. Journal of Neuroscience, 27(40), 10659-10673.

Alcock, J. (2001). Animal behavior: an evolutionary approach. Sinauer Associates, 7ed.

Aldridge, J. W., Berridge, K. C., Herman, M., & Zimmer, L. (1993). Neuronal coding of serial order: Syntax of grooming in the neostriatum. Psychological Science, 4, 391–395.

Alexander, G. E, DeLong, M. R, Strick, P. L (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annual Review of Neuroscience.

Alexander, G. E., Crutcher, M. D. (1990). Functional architecture of basal ganglia circuits: neural substrates of parallel processing. TINS, 13(7), 266-271.

Allen, G. I., Tsukahara, N. (1974). Cerebrocerebellar communication systems. Physiological Reviews, 54(4), 957-1006.

Andersen, B. B., Korbo, L., and Pakkenberg, B. (1992). A quantitative study of the human cerebellum with unbiased stereological techniques. J. Comp. Neurol. 326, 549–560.

Ashe, J., Lungu, O. V., Basford, A. T., Lu, X. (2006). Cortical control of motor sequences. Current Opinion in Neurobiology, 16, 213-221.

Ashby, F. G., Turner, B. O., Horvitz, J. C. (2010). Cortical and basal ganglia contributions to habit learning and automaticity. Trends in Cognitive Sciences, 14(5), 208-215.

Averbeck, B. B., Chafee, M. V., Crowe, D. A., Georgopoulos, A. P. (2002). Parallel processing of serial movements in prefrontal cortex. PNAS, 99(20), 13172-13177.

Averbeck, B. B., Sohn, J., & Lee, D. (2005). Activity in prefrontal cortex during dynamic selection of action sequences. Nature Neuroscience, 9, 276–282.

Angulo-Kinzler, R. (2001). Exploration and selection of intralimb coordination patterns in 3-month-old infants. Journal of Motor Behavior, 33(4), 363-376.

Baillargeon, R., Spelke, E. S., Wasserman, S. (1985). Object permanence in 5-month-old

infants. Cognition, 20, 191-208.

Bapi, R. S, Doya, K, Harner, A. M (2000). Evidence for effector independent and dependent representations and their differential time course of acquisition during motor sequence learning. Experimental Brain Research, 132, 149-62.

Bapi, R. S., Doya, K. Multiple forward model architecture for sequence processing. In R. Sun & L. Gilles (Eds.). Sequence learning: paradigms, algorithms, and applications. Springer, 2001.

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. TRENDS in Cognitive Sciences, 11(7), 280-289.

Barnes, T. D., Kubota, Y., Hu, D., Dezhe, Z. J., Graybiel, A. M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of precedural memories. Science, 437, 1158-1161.

Barnett, S. M., Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. Psychological Bulletin, 128(4), 612-637.

Bartels, A., Zeki, S. (2000). The neural basis of romantic love. Neuroreport, 11, 3829-3834.

Bartels, A., Zeki, S. (2004). The neural basis of maternal and romantic love. Neuroimage,

21, 1155-1166.

Barton, R. A. (2006). Primate brain evolution: integrating comparative, neurophysiological, and ethological data. Evolutionary Anthropology, 15, 224-236.

Barton, R. A., and Harvey, P. H. (2000). Mosaic evolution of brain structure in mammals. Nature, 405, 1055–1058.

Beer, J. S., John, O. P., Scabini, D., Knight, R. T. (2006). Orbitofrontal cortex and social behavior: integrating self-monitoring and emotion-cognition interactions. Journal of Cognitive Neuroscience, 18(6), 871-879.

Bekoff, M. (2002). The cognitive animal: empirical and theoretical perspectives on animal cognition. MIT Press.

Blakemore, S-J., Wolpert, D. M., Frith, C. D. (2002). Abnormalities in the awareness of action. TINS, 6(6), 237-242.

Bostan, A. C., Dum, R. P., Strick, P. L. (2010). The basal ganglia communicate with the cerebellum. PNAS, 107(18), 8452-8456.

Bralet, M. C., Navarre, M., Eskenazi, A. M., Lucas-Ross, M., Falissard, B. (2008). Interest of a new instrument to assess cognition in schizophrenia: the Brief Assessment of Cognition in Schizophrenia (BACS). Encephale, 34(6), 557-562.

Bruce. D. (1994). Lashley and the problem of serial order. American Psychologist, 49(2), 93-103.

Burish M. J., Peebles, J. K., Baldwin, M. K. , Tavares, L., Kaas, J. H., Herculano-Houzel, S.(2010). Cellular scaling rules for primate spinal cords. Brain, Behavior and Evolution, 76(1), 45-59.

Carpenter, A. F., Georgopoulos, A. P., Pellizzer, G. (1999). Motor cortical encoding of serial order in a context-recall task. Science, 283, 1752-1757.

Carr, E. G., Durand, V. M. (1985). Reducing behavior problems through functional communication training. Journal of Applied Behavioral Analysis, 18(2), 111-126.

Clegg, B. A, DiGirolano, G. J, Keele, S. W (1998) Sequence learning. Trends in Cognitive Sciences, 2(8): 275-281.

Daw, N. D, Niv, Y, Dayan, P (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nature Neuroscience, 8(12):1704-11.

Doane, S. M., Alderton, D. L., Sohn, Y. W., Pellegrino, J. W. (1995). Acquisition and transfer of skilled performance: are visual discrimination skills stimulus specific? Journal of Experimental Psychology: human perception and performance, 22(5), 1218-1248.

Dommett, E. et al. (2005). How visual stimuli activate dopaminergic neurons at short latency. Science, 307, 1476-1480.

Doya, K., Sejnowski, T. J. (1999). A computational model of avian song learning. In Gazzaniga M. S., The New Cognitive Neurosciences, MIT Press, 469-482.

Doya, K (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex. Neural Networks, 12, 961-974.

Doya, K (2000). Complementary roles of basal ganglia and cerebellum in learning and motor control. Current Opinion in Neurobiology, 10:732-739.

Doya, K (2007). Reinforcement Learning: Computational theory and biological mechanisms. HFSP, 1:30-40.

Doyon, J., Owen, A.M., Petrides, M., Sziklas, V., Evans, A.C. (1996).Functional anatomy of visuomotor skill learning in human subjects examined with positron emission tomography. European Journal of Neuroscience, 8, 637–648.

Doyon, J., Laforce Jr., R., Bouchard, G., Gaudreau, D., Roy, J., Poirier,M., Bedard, P.J., Bedard, F., Bouchard, J.P., 1998. Role of the striatum, cerebellum and frontal lobes in the in the automatization of a repeated visuomotor sequence of movements. Neuropsychologia 36, 625–641.

Doyon J., Song, A. W., Karni A., Lalonde, F., Adams, M. M,, Ungerleider, L. G. (2002). Experience dependent changes in cerebellar contributions to motor sequence learning. PNAS, 99, 1017–22.

Dow, R. S., Moruzzi, G. (1958). The physiology and pathology of the cerebellum. Minneapolis: University of Minnesota Press.

Dum, R. P., Li, C., Strick, P. L. (2002). Motor and nonmotor domains in the monkey dentate. Annals of the New York Academy of Sciences, 978, 289-301.

Dum, R. P., Strick, P. L. (2003). An unfolded map of the cerebellar dentate nucleus and its projections to the cerebral cortex. Journal of Neurophysiology, 89(1), 634-639.

Duncan, C. P. (1953). Transfer in motor learning as a function of degree of first-task learning and inter-task similarity. Journal of Experimental Psychology, 45(1), 1-11.

Eblen, F., Graybiel, A. M. (1995). Highly restricted origin of prefrontal cortical inputs to striosomes in the macaque monkey. Journal of Neuroscience, 15, 5999-6013.

Félix, M. A., M; Braendle, C. C. (2010). The natural history of Caenorhabditis elegans. Current Biology, **20** (22): R965–R969.

Fisher, H. E., Aron, A., Brown, L. L. (2006). Romantic love: a mammalian brain system for mate choice. Philosophical Transactions of the Royal Society B, 361, 2173-2186.

Fitts, P. M. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), Categories of human learning (pp. 243-285). New York: Academic Press.

Fitts, P. M. & Posner, M. I (1967). Human Performance. Belmont, CA: Brooks/Cole.

Fogel, A., Thelen, E. (1987). Development of early expressive and communicative action: reinterpreting the evidence from a dynamic systems perspective. Developmental Psychology, 23(6), 747-761.

Flaherty, A. W., Graybiel, A. M. (1993). Output architecture of the primate putamen. Journal of Neuroscience, 13, 3222-3237.

Flanagan, J. R., Vetter, P., Johansson, R. S., Wolpert, D. M. (2003). Prediction precedes control in motor learning. Current Biology, 13, 146-150.

Floyer-Lea, A., Matthews, P. M. (2005). Distinguishable brain activation networks for short-term motor skill learning. Journal of Neurophysiology, 94, 512-518.

Frankle, W. G., Laruelle, M., Haber, S. N. (2006). Prefrontal cortical projections to the midbrain in primates: evidence for a sparse connection. Neuropsychopharmacology, 31, 1627-1636.

Frith, C. D., Blakemore, S-J., Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. Philosophical Transactions of the Royal Society of London B, 355, 1771-1788.

Fudge, J. L., Haber, S. N. (2000). The central nucleus of the amygdala projection to dopamine subpopulations in primates. Neuroscience, 97, 479-497.

Fujii, N., Graybiel, A. M. (2003). Representation of action sequence boundaries by macaque prefrontal cortical neurons. Science, 301(5637), 1246-1249.

Funahashi, S., Goldman-Rakic, P.S., and Bruce, C.J. (1992) Mnemonic coding of visual space in the primate prefrontal cortex revealed by oculomotor paragigms. In: "Perspectives in Neuroethology," K. Kubota (ed.), Kyoto Univ. Press, Kyoto, Japan, pp. 137-151.

Funahashi, S., Bruce, C.J., and Goldman-Rakic, P.S. (1991) Neuronal activity related to saccadic eye movements in the monkey's dorsolateral prefrontal cortex.? Journal of Neurophysiology, 65: 1464-1483.

Gentile, A. M. (1972). A working model of skill acquisition with application to teaching. Quest Monograph XVII, 3-23.

Gesell, A. L. (1939). Reciprocal interweaving in neuromotor development. Journal of Comparative Neurology, 70, 161-180.

Ghilardi, M.F., Alberoni, M., Marelli, S., Rossi, M., Franceschi, M., Ghez, C., Fazio, F. (1999). Impaired movement control in Alzheimer's disease. Neuroscience Letters, 260, 45–48.

Gibson, E. J. (1969). Principles of perceptual learning and development. New York: Academic.

Gibson, J. J. (1979). The ecological approach to visual perception. New York: Houghton Mifflin.

Glanzer, M. (1958). Curiosity, exploratory drive, and stimulus satiation. Psychological Bulletin, 55(5), 302-315.

Glascher, J., Daw, N. D., Dayan, P., O'doherty, J. P. (2010). States vs rewards: dissociable neural prediction error signals underlying mdoel-based and model-free reinforcement learning . Neuron 66, 585-595.

Gläscher, J., Hampton, A. N., & O'Doherty, J. P. (2009). Determining a Role for Ventromedial Prefrontal Cortex in Encoding Action-Based Value Signals During Reward-Related Decision Making. Cerebral Cortex, 19(2), 483 -495.

Glencross, D. J (1977). Control of skilled movements. Psychological Bulletin, 84(1): 14-29.

Glickstein, M., Doron, K. (2008). Cerebellum: connections and functions. Cerebellum, 7, 589-594.

Goldfield, E. C. (1989). Transition from rocking to crawling: postural constraints on infant movement. Developmental Psychology, 25(6), 913-919.

Goldman-Rakic, P.S., Funahashi, S. and Bruce, C.J. (1990) Neocortical memory Circuit. Cold Spring Harbor Symposia on Quantitative Biology, 55: 1025-1038.

Goldman-Rakic, P. S (1995). Toward a circuit model of working memory and the guidance of voluntary motor action. IN: Houk et. al (1995) Models of information processing in the basal ganglia. MIT press.

Grafton, S. T., Hazeltine, E., Ivry, R. B. (1997). Functional mapping of sequence learning in normal humans. Journal of Neuroscience, 7, 497-510.

Graybiel, A. M., Aosaki, T., Flaherty, A. W., Kimura, M. (1994). The basal ganglia and adaptive motor control. Science, 265, 1826-1831.

Gusnard, D. A., Akbudak, E., Shulman, G., L., Raichle, M. E. Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. PNAS, 98(7), 4259-4264.

Haber, S. N., Fudge, J. L., McFarland, N. R. (2000). Striatonigral pathways in primates

form an ascending spiral from the shell to the dorsolateral striatum. Journal of Neuroscience, 20, 2369-2382.

Haber, S. N. The primate basal ganglia: parallel and integrative networks. Journal of Chemical Neuroanatomy, 26, 317-330, 2003.

Haber, S. N., Kim, K. S., Mailly, P., Calzavara, R. (2006). Reward-related cortical inputs define a large striatal region in primates that interface with associative cortical inputs, providing a substrate for incentive-base learning. Journal of Neuroscience, 267, 8368-8376.

Hajime, M., Saito, N., Furusawa, Y., Izumiyama, M., Sakamoto, K., Shamoto, H., Shimizu, H., Yoshimoto, T. (2002). Orderly activations of human cortical areas during path-planning task. Neuroreport, 13, 423-426.

Halsband, U., Lange, R. K. (2006). Motor learning in man: a review of functional and clinical studies. Journal of Physiology, 99, 414-424.

Haug, H. (1987). Brain sizes, surfaces, and neuronal sizes of the cortex cerebri: a stereological investigation of man and his variability and a comparison with some mammals (primates, whales, marsupials, insectivores, and one elephant). Am. J. Anat. 180, 126–142.

Herculano-Houzel, S., and Lent, R. (2005). Isotropic fractionator: a simple, rapid method

for the quantification of total cell and neuron numbers in the brain. Journal of Neuroscience, 25, 2518–2521.

Herculano-Houzel, S. (2009). The human brain in numbers: a linearly scaled-up primate brain. Frontiers in Human Neuroscience, 3(31).

Hikosaka, O., Rand, M. K., Miyachi, S., Miyashita, K. Learning of sequential movements in the monkey: process of learning and retention of memory. Journal of Neurophysiology, 74, 1652-61, 1995.

Hikosaka, O, Nakahara, H, Rand, M. K, Sakai, K, Lu, X, Nakamura, K, Miyachi, S, Doya, K (1999). Parallel neural networks for learning sequential procedures. Trends in Neurosciences, 22: 464-471.

Hikosaka, O, Nakamura, K, Sakai, K, Nakahara, H (2002). Central mechanisms of motor skill learning, Current Opinion in Neurobiology, 12(2): 6315-20.

Hofsten, C. von. (1984). Developmental changes in the organization of prereaching movements. Developmental Psychology, 20, 378-388.

Horvitz, J. C. (2009). Stimulus-response and respons-outcome learning mechanisms in the striatum. Behavioral Brain Research, 199, 129-140.

Hoshi, E., Tremblay, L., Feger, J., Carras, P. L., Strick, P. L. (2005). The cerebellum

communicates with the basal ganglia. Nature Neuroscience, 8, 1491-1493.

Houk, J. C., Adams, J. L., Barto, A. C. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In Models of Information Processing in the Basal Ganglia. J. C. Houk, J. L. Davis, D. G. Beiser, eds., The MIT Press, 1995, pp. 249-270.

Hutton, S. B., Puri, B. K., Duncan, L. J., Robbins, T. W., Barnes, T. R., Joyce, E. M. (1998). Executive function in first-episode schizophrenia, Psychological Medicine, 28(2), 463-473.

Ito, M (2008). Control of mental activities by internal models in the cerebellum. Nature Neuroscience, 23:49-56.

Ito, M., Doya, K. (2009). Validation of decision-making models and analysis of decision variables in the rat basal ganglia. Journal of Neuroscience, 29(31), 9861-9874.

James, W. (1890). Principles of Psychology. Vol I. New York: Holt.

James, W. (1914). Habit. New York: Holt.

Jenkins, I. H., Brooks, D. J., Nixon, P. D., Frackowiak, R. S. J., Passingham, R. E. (1994). Motor sequence learning: a study with positron emission tomography. Journal of Neuroscience, 14, 3775-3790

Johnson-Laird, P. N (1983). Mental Models: towards a cognitive science of language, inference, and consciousness. Cambridge: Cambridge University Press; Cambridge, MA: Harvard University Press.

Jubault, T., Ody, C., Koechlin, E. (2007). Serial organization of human behavior in the inferior parietal cortex. Journal of Neuroscience, 27(41), 11028-11036.

Judd, C. H (1908). The relation of special training to general intelligence. Educational Review, 36: 28-42.

Jueptner, M., Stephan, K. M., Frith, C. D., Brooks, D. J., Frackowiak, R. S. J., Passingham, R. E. (1997a). Anatomy of motor learning. I. Frontal cortex and attention to action. Journal of Neurophysiology, 77, 1313-1324.

Jueptner, M., Frith, C. D., Brooks, D. J., Frackowiak, R. S. J., Passingham, R. E. (1997b). Anatomy of motor learning. II. Subcortical structures and learning by trial and error. Journal of Neurophysiology, 77, 1325-1337.

Kanner, L., Rodriguez, A., Ashenden, B. (1972). How far can autistic children go in matters of social adaptation? Journal of Autism Child and Schizophrenia, 2(1), 9-33.

Kawato, M (1999). Internal models for motor control and trajectory planning. Current Opinion in Neurobiology, 9:718-727.

Keele, S, Ivry, R, Mayr, U, Hazeltine, E, Heuer, H (2003). The cognitive and neural architecture of sequence representation. Psychological Review, 110:316-339.

Kelly, R. M., Strick, P. L. (2003). Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. Journal of Neuroscience, 23, 8432-8444.

Kennerley, S. W., Sakai, K., & Rushworth, M. F. S. (2004). Organization of action sequences and the role of the pre-SMA. Journal of Neurophysiology, 91, 978–993.

Koechlin, E., Jubault, T. (2006). Broca area and the hierarchical organization of human behavior. Neuron, 50(6), 963-974.

Koegel, R. L., Koegel, L. K.(1990). Extended reductions in stereotypic behavior of students with autism through a self-management treatment package. Journal of Applied Behavioral Analysis, 23(1), 119-127.

Koegel, L. K., Koegel, R. L., Hurley, C., Frea, W. D. (1992). Improving social skills and disruptive behavior in children with autism through self-management. Journal of Applied Behavioral Analysis, 25(2), 341-353.

Konczak, J., Dichgans, J. (1997). The development toward stereotypic arm kinematics during reaching in the first 3 years of life. Experimental Brain Research, 117, 346-354.

Lange, W. (1975). Cell number and cell density in the cerebellar cortex of man and some

other mammals. Cell Tissue Res. 157, 115–125.

Lashley, K. S. (1917). The accuracy of movement in the absence of excitation from the moving organ. American Journal of Physiology, 43, 169-194.

Lashley, K. S. (1948). The mechanism of vision. XVIII. Effects of destroying the visual associative areas in the monkey. Genetic Psychology Monographs, 37, 107-166.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), Cerebral mechanisms in behavior (pp. 112–131). New York: Wiley.

Lee, T. D., Magill, R. A. (1983). The locus of contextual interference in motor-skill acquisition. Journal of Experimental Psychology: learning, memory and cognition, 9, 730-746.

Lee, T. D (1988). Testing for motor learning: a focus on transfer-appropriate processing. In O. G. Meijer & K. Roth (Eds.), Complex motor behavior: The motor-action controvery (p. 210-215). Amsterdam: Elsevier.

Lee, J-Y., Schweighofer. Dual adaptation supports a parallel architecture of motor memory. The Journal of Neuroscience, 29, 10396-404, 2009.

Leherecy, S et al. (2005). Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. PNAS, 102(35), 12566-12571.

Lewandowsky, S., Murdoch, B. B. Jr. (1989). Memory for serial order. Psychological Review, 96(1), 25-57.

Ljungberg, T., Apicella, P., Schultz, W. (1992). Responses of monkey dopamine neurons during learning of behavioral reactions. Journal of Neurophysiology, 67(1), 145-163.

Llinas, R., Hillman, D. E. (1969). Physiological and morphological organization of the cerebellar circuits in various vertebrates. In R. Llinas (Ed), Neurobiology of cerebellar evolution and development (pp. 43-73). Chicago: American Medical Association.

Lu, X., Hikosaka, O., Miyachi, S. (1998). Role of monkey cerebellar nuclei in skill for sequential movement. Journal of Neurophysiology, 79, 2245-2254.

Lynd-Balta, E., Haber, S. N. (1994a). Primate striatonigral projections: a comparison of the sensorimotor-related striatum and the ventral striatum. Journal of Comparative Neurology, 345, 562-578.

Lynd-Balta, E., Haber, S. N. (1994b). The organization of midbrain projections to the striatum in the primate: sensorimotor-related striatum vs ventral striatum. Neuroscience, 59, 625-604.

Lynd-Balta, E., Haber, S. N. (1994c). The organization of midbrain projections to the ventral striatum in the primate. Neuroscience, 59, 609-623.

McCulloch, W. S. (1945). A heterarchy of values determined by the topology of nervous nets. Bulletin of Mathematical Biophysics, 7, 89-93.

McGeoch, J. A. (1931). The acquisition of skill. Psychological Bulletin, 28(6), 413-466.

McGraw, M. G. (1932). From reflex to muscular control in the assumption of an erect posture and ambulation in the human infant. Child Development, 3, 291-297.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. Science, 198, 75-78.

Middleton, F. A., Strick, P. L. (2000). Basal ganglia and cerebellar loops: motor and cognitive circuits. Brain Research Reviews, 31, 236-250.

Molinari, M., Leggio, M. G., Solidy, A., Clorra, R., Misciagna, S., Silveri, M. C., Pertrosoni, L. (1997). Cerebellum and procedural learning: evidence from cerebellar lesions. Brain, 120, 1753-1762.

Miyachi, S., Hikosaka, O., Miyashita, K., Karadi, Z., Rand, M. K. (1997). Differential roles of monkey striatum in learning of sequential hand movement. Experimental Brain Research, 115, 1-5.

Murakami, Y., Tanaka, M. (2011). Evolution of motor innervation to vertebrate fins and limbs. Developmental Biology, 355(1), 164-172.

Mushiake H, Saito N, Sakamoto K, Itoyama Y, Tanji J. (2006). Activity in the lateral prefrontal cortex reflects multiple steps of future events in action plans. Neuron, 50(4), 631-41.

Mushiake H, Sakamoto K, Saito N, Inui T, Aihara K, Tanji J. (2009). Involvement of the prefrontal cortex in problem solving. International Review of Neurobiology, 85, 1-11.

Muslimovic, D., Post, B., Speelman, J. D., Schmand, B. (2007). Motor procedural learning in Parkinson's disease. Brain, 130(11), 2887-2897.

Nakamura, K., Sakai, K., Hikosaka, O. (1998). Neuronal activity in medial frontal cortex during learning of sequential procedures. Journal of Neurophysiology, 80, 2671-2687.

Nakano, K., Kayahara, T., Tsutsumi, T., Ushiro, H. (2000). Neural circuits and functional organization of the striatum. Journal of Neurology, 247(S5), V/1-V5.

Nakahara, H, Doya, K, Hikosaka, O (2001). Parallel cortico-basal ganglia mechanisms for acquisition and execution of visuo-motor sequences – A computational approach. Journal of Cognitive Neuroscience, 13:626-647.

Newell, K (1985). Coordination, control and skill. In D. Goodman, R. B. Wilberg & I. M. Franks (Eds.), Differing perspectives in motor learning, memory, and control (p. 299-317. Amsterdam: North-Holland.

Ninokura Y, Mushiake H, Tanji J.(2003). Representation of the temporal order of visual objects in the primate lateral prefrontal cortex. Journal of Neurophysiology, 89(5), 2868-73.

Passingham, R. E. Attention to action. Philosophical Transactions of the Royal Society of London B, 351, 1473-49, 1996.

Patniyot, N. S. (2011). Thought disorder in schizophrenia: impairment in contextual processing via integrative failures in cognition. Medical Hypotheses (in press).

Rand, M. K., Hikosaka, O., Miyachi, S., Lu, X., Miyashita, K. (1998). Characteristics of a long-term procedural skill in the moneky. Experimental Brain Research, 118, 293-297.

Rand, M. K., Hikosaka, O., Miyachi, S., Lu, X., Nakamura, K., Kitaguchi, K., Shimo, Y. (2000). Characteristics of sequential movements during early learning period in monkeys. Experimental Brain Research, 131, 293-304.

Ramnani, N. (2006). The primate cortico-cerebellar system: anatomy and function. Nature Neuroscience, 7, 511–522.

Redgrave, P., Gurney, K. (2006). The short-latency dopamine signal: a role in discovering novel actions? Nature Reviews Neuroscience, 7, 967-975.

Rochat, P., Goubet, N. (1995). Development of sitting and reaching in 5- to 6-month-old

infants. Infant Behavior and Development, 18, 53-68

Rosenbaum, D. A., Kenny, S., & Derr, M. A. (1983). Hierarchical control of rapid movement sequences. Journal of Experimental Psychology: Human Perception and Performance, 9, 86–102.

Rosenbaum, D. A., Carlson, R. A., & Gilmore, R. O. (2001). Acquisition of intellectual and perceptual-motor skills. Annual Review of Psychology, 52, 453–470.

Rosenbaum, D. A., Cohen, R. G., Meulenbroek, R. G., & Vaughan, J. (2006). Plans for grasping objects. In M. Latash & F. Lestienne (Eds.), Motor control and learning over the lifespan (pp. 9–25). Berlin: Springer.

Rosenbaum, D. A., Cohen, R. G., Jax, S. A., Weiss, D. J., van der Wel, R. (2007). The problem of serial order in behavior: Lashley's legacy. Human Movement Science, 26, 525-554.

Rauch, S.L., Savage, C.R., Brown, H.D., Curran, T., Alpert, N.M.,Kendrick, A., Fischman, A.J., Kosslyn, S.M., 1995. A PET investigationof implicit and explicit sequence learning. Hum. Brain Mapp. 3, 271–286.

Rauch, S.L., Whalen, P.J., Savage, C.R., Curran, T., Kendrick, A., Brown, H.D., Bush, G., Breiter, H.C., Rosen, B.R., 1997. Striatal recruitment during an implicit sequence learning task as measured by functional magnetic resonance imaging. Hum. Brain Mapp.

5, 124–132.

Roth, G., and Dicke, U. (2005). Evolution of the brain and intelligence. Trends Cogn. Sci. 9, 250–257.

Saito N, Mushiake H, Sakamoto K, Itoyama Y, Tanji J. (2005). Representation of immediate and final behavioral goals in the monkey prefrontal cortex during an instructed delay period. Cerebral Cortex, 15(10):1535-46.

Sakai, K., Hikosaka, O., Miyauchi, S., Takino, R., Sasaki, Y., Putz, B. (1998). Transition of brain activity from frontal to parietal areas in visuomotor sequence learning. Journal of Neuroscience, 15(5), 1827-1840.

Sakai, K., Kitaguchi, K., Hikosaka, O. (2003). Chunking during human visuomotor sequence learning. Experimental Brain Research, 152, 229-242.

Sakai, K, Hikosaka, O, Nakamura, K (2004). Emergence of rhythm during motor learning. Trends in Cognitive Sciences, 8(12): 547-53.

Samejima, K., Doya, K. (2007). Multiple representations of belief states and action values in cortico-basal ganglia loops. Annals of the New York Academy of Science.

Schack, T & Mechsner, F (2006). Representation of motor skills in human long-term memories. Neuroscience Letters, 391: 77-81.

Schimidt, R. A. (1975). A schema theory of discrete motor skill learning. Psychological Review, 82, 225-260.

Seidler, R (2004). Multiple motor learning experiences enhance adaptability. Journal of Cognitive Neuroscience, 16: 65-73.

Seidler, R (2005). Differential transfer processes in incremental visuomotor adaptation. Motor Control, 9: 40-58.

Selemon, L. D., Goldman-Rakic, P. S. (1985). Longitudinal topography and interdigitation of corticostriatal projections in the rhesus monkey. Journal of Neuroscience, 5, 776-794.

Selemon, L. D., Goldman-Rakic, P. S. (1988). Common cortical and subcortical targets of the dorsolateral prefrontal and posterior parietal cortices in the rhesus monkey: evidence for a distributed neural networks subserving spatially guided behavior. Journal of Neuroscience, 8, 4049-4968.

Seo, H., Barraclough, D. J., Lee, D. (2009). Lateral intraparietal cortex and reinforcement learning during a mixed-strategy game. Journal of Neuroscience, 29(22), 7278-7289.

Schneider, W., Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. Psychological Review, 84(1), 1-66.

Shadmehr, R, Mussa-Ivaldi, F.A (1994). Adaptive representation of dynamics in human motor learning. Journal of Neuroscience, 14:3208-3224.

Shadmehr, R., Krakauer, J. W. (2008). A computational neuroanatomy for motor control. Experimental Brain Research, 185, 359-381.

Shah, A, Barto, A. G. Effect on movement selection of an evolving sensory representation: a multiple controller model of skill acquisition. Brain Research, 1299, 55-73, 2009

Shea, J. B., Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. Journal of Experimental Psychology: human learning and memory, 5(2), 179-187.

Sherrington, C. S. (1906). The integrative action of the nervous system. Yale University Press.

Shiffrin, R. M., Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. Psychological Review, 84(2), 127-190.

Shima K, Isoda M, Mushiake H, Tanji J. (2007). Categorization of behavioural sequences in the prefrontal cortex. Nature, 445(7125), 315-8.

Skinner, B. F. (1975). The shaping of phylogenic behavior. Acta Neurobiologiae

Experimentalis, 35(5), 409-415.

Skinner, B. F. The evolution of behavior (1984). Journal of the Experimental Analysis of Behavior, 41(2), 217-221.

Sheridan, M. R (1984). Response programming, response production, and fractionated reaction time. Psychological Research, 46:33-47.

Simon, D. A., Daw, N. D. (2011). Neural correlates of forward planning in a spatial decision task in humans**. J**ournal of neuroscience, 31, 5526-39.

Smith, M. A., Ghazizadeh, A., Shadmehr, R. Interacting adaptive processes with different timescales underlie short-term motor learning. Plos Biology, 4, 1-9, 2006.

Spelke, E. S., Breinlinger, K., Macomber, J., Jacobson, K. (1992). Origins of knowledge. Psychological Review, 106, 235-260.

Sugden, D. A., Chambers, M. E. (2005). Children with developmental coordination disorder. London: Whurr Publishers Ltd.

Sultan, F. (2002). Analysis of mammalian brain architecture. Nature 415, 133–134.

Sutton, R. S & Barto, A. G (1998). Reinforcement Learning: An Introduction. Cambridge, MA, MIT press.

Suvorov, N. F., Shuvaev, V. T., Voilokova, N. L., Chivileva, O. G., Shefer, V. I. (1997). Corticostriatal mechanisms of behavior. Neuroscience and Behavioral Psychology, 27(6), 653-662.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), Information Processing in Motor Control and Learning (pp. 117–152). New York: Academic Press.

Szpunar, K. K., Watson, J. M., McDermott, K. B. (2007). Neural substrates of envisioning the future. PNAS, 104(2), 642-647.

Tanji, J., Shima, K. (1994). Role for supplementary motor area cells in planning several movements ahead. Nature, 371, 413-416.

Tanji, J., Shima, K., Mushiake, H. (1996). Multiple cortical motor areas and temporal sequencing of movements. Cognitive Brain Research, 5, 117-122.

Tanji, J., Hoshi, E., (2001). Behavioral planning in the prefrontal cortex. Current Opinion in Neurobiology, 11, 164-170.

Tanji, J. (2001). Sequential organization of multiple movements: involvement of cortical motor areas. Annual Review of Neuroscience, 24, 631-651.

Taylor, M. E. (2009). Trahsfer in reinforcement learning domains. Berlin: Springer.

Thelen, E., Corbetta, D., Kamm, K., Spencer, J. P., Schneider, K., Zernicke, R. F. (1993). The transition to reaching: mapping intention and intrinsic dynamics. Child Development, 64(4), 1058-1098.

Thelen, E., Smith, L. B. (1994). A dynamic systems approach to the development of cognition and action. Cambridge, MA: Bradford Books/MIT Press.

Thelen, E., Schoner, G., Scheier, C., Smith, L. B. (2001). The dynamics of embodiment: a field theory of infant perseverative reaching. Behavioral and Brain Sciences, 24, 1-86.

Thorndike, E. L. (1911). Animal intelligence. Darien, Conn: Hafner.

Thorndike, E. L (1927). The law of effect. American Journal of Psychology, 29: 212-222.

Thorndike, E. L. & Woodworth, R. S (1901). The influence of improvement in one mental function upon the efficiency of other functions (I). Psychological Review, 8: 247-261.

Todorov, E. (2004). Optimality principles in sensorimotor control. Nature Neuroscience, 7, 907-915.

Toni, I., Krams, M., Turner, R., Passingham, R. E. (1998). The time course of changes

during motor sequence learning: a whole-brain fMRI study. Neuroimage, 8, 50-61.

Turvey, M (1991). Coordination. American Psychologist, 45: 938-953.

Vereijken, B, Whiting, H. T. A & Beek, P. J (1992). A dynamical systems approach to skill acquisition. Quarterly Journal of Experimental Psychology, 45A: 323-344.

Verwey, W. B. (1995). A forthcoming key press can be selected while earlier ones are executed. Journal of Motor Behavior, 27(3), 275-284.

Wang, J & Sainburg, R. L (2003). Mechanisms underlying interlimb transfer of visuomotor rotations. Experimental Brain Research, 149: 520-526.

Watkins, C. J. C. H., Dayan, P. Q-Learning. Machine Learning, 8, 279-92, 1992.

Whiting, B. A., and Barton, R. A. (2003). The evolution of the cortico-cerebellar complex in primates: anatomical connections predict patterns of correlated evolution. J. Human Evolution. 44, 3–10.

Willingham, D. B. (1998). A neuropsychological theory of motor skill learning. Psychological Review, 105(3), 558-584.

Xin, J., Costa, R. M. (2010). Start/stop signals emerge in nigrostriatal circuits during sequence learning. Nature, 466, 457-462.

Wolpert, D. M., Kawato, M. (1998). Multiple paired forward and inverse models for motor control. Neural Networks, 11, 1317-1329.

Wolpert, D. M., Flanagan, J. R. (2010). Motor learning. Current Biology, 20(11), R467-472.

Wood, J. N. (2006). Social cognition and the prefrontal cortex. Behavioral and Cognitive Neuroscience Reviews, 5(4)

Woodward, P. (1943). An experimental study of transfer of training in motor learning. Journal of Applied Psychology, 27(1), 12-32.

Wunderlich, K., Rangel, A., O'Doherty, J.P. (2009). Neural computations underlying action-based decision making in the human brain, Proc Natl Acad Sci USA. 106(40):17199-17204.

Yoshida, W., Ishii, S. (2006). Resolution of uncertainty in prefrontal cortex. Neuron, 50, 781-789.