An Empirical and Computational Investigation of Skill Learning in Internally-guided Sequencing

Thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science
in
Exact Humanities by Research

by

Krishn Bera
201556009
krishn.bera@research.iiit.ac.in

International Institute of Information Technology
Hyderabad - 500 032, INDIA
June 2021

Copyright © Krishn Bera, 2021
All Rights Reserved

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled **"An Empirical and Computational Investigation of Skill Learning in Internally-guided Sequencing"** by **Krishn Bera**, has been carried out under my supervision and is not submitted elsewhere for a degree.

_____ _____

Date Adviser: Prof. Bapi Raju S.

To, *Professor Navjyoti Singh*

Acknowledgments

First and foremost, I would like to thank my advisor, Prof. Bapi Raju, whose mentorship and guidance have been very formative for me as a student and a researcher in cognitive science. I thank him for providing me with an opportunity to explore my interests and delve deeper into the domain. This enterprise would not have been possible without him and I thank him for motivating my interest in pursuing the domain further. I also express my gratitude towards Late Prof. Navjyoti Singh; his teachings, ideas, and conversations inspire me to pursue interdisciplinary research at the frontiers of confluence domains.

I would like to thank my friend and mentor, Anuj Shukla. This work has immensely benefited from the countless stimulating discussions that were a constant source of advice, ideas and encouragement for me. I also thank Madhukar, Aditya and Shivang, my fellow companions and wanderers in (almost) everything. I thank my friends - Chaitanya, Kishan, Kulin, Tirth and Gopal, for being an invaluable part of my journey and a million fond memories at IIIT-H and beyond. I would also like to thank my CHD crew for making a great company throughout the years.

Finally, I would like to thank my parents for their unwavering support and belief in me. None of this would have been possible without them.

Abstract

Sequence learning plays a central role in the acquisition of many daily life motor skills such as typing or playing the piano. Several canonical experimental paradigms such as the serial reaction time task, discrete sequence production task and $m \times n$ task have been proposed to study the typical behavioral phenomenon in sequencing tasks. Such paradigms are externally-specified, where the environment or the task paradigm extrinsically provides the sequence of stimuli that guides the motor actions. Such paradigms differ from a class of more realistic motor tasks that are internally-guided, where the sequence of motor actions is self-generated or internally-specified. Most previous studies on discrete sequencing have employed externally-specified paradigms and therefore, the cognitive mechanisms underlying skill learning in internally-guided sequencing paradigms remain largely unexplored.

This thesis presents an empirical and computational investigation of skill learning in internally-guided sequencing. We employ the Grid-Sailing Task (GST) as a canonical paradigm to study internally-guided sequence learning. The GST requires navigating by executing sequential keypresses, a $n \times n$ grid from start to goal (SG) position while using a particular key-mapping (KM) among the three cursor-movement directions and the three keyboard buttons.

In the first study, we investigate the learning processes involved in internally-guided sequencing. The participants performed two behavioral experiments – Single-SG and Mixed-SG condition. The participants first completed the Single-SG condition, which required performing GST on a single SG position repeatedly. By showing performance-related improvements in various behavioral measures such as the execution time and reward score, we show that motor learning contributes to the trajectory-specific learning in GST with the repeated execution of the same keypress sequences. The Mixed-SG condition involved performing GST using the same KM (from Single-SG condition) on two novel SG positions presented in a random, inter-mixed manner. Since the participants utilize the previously learned KM, we anticipate a transfer of learning from the Single-SG condition. The acquisition and transfer of a KM-specific internal model facilitate efficient trajectory planning on novel SG conditions. The acquisition of such a KM-specific internal model amounts to trajectory-independent cognitive learning in GST. We provide evidence for the role of cognitive learning in GST by showing transfer-related performance improvements in the Mixed-SG condition.

In a subsequent study, we probe the involvement of a particular motor learning process called motor chunking. Motor chunking is a phenomenon which enables efficient execution of the motor sequences by chaining several elementary actions into sub-sequences called motor chunks. The participants per-

formed GST on a 10×10 grid, executing the same trajectory repeatedly throughout the experiment. We provide empirical evidence for motor chunking by showing the emergence of subject-specific, unique temporal patterns in response times. Our findings show spontaneous chunking without pre-specified or externally guided structures while replicating the earlier results with a less constrained, internally guided sequencing paradigm.

In another study, we employ an inter-manual transfer task to examine the stage-wise transitions in motor sequence learning. The participants performed GST on inter-leaved normal and transfer blocks. The dominant hand was used on the normal block and the non-dominant hand was used on the transfer block. The length of the first normal block varied across days. We found increasing differences in execution time between the normal and transfer blocks across days as the effector-dependent learning consolidated. Our findings confirm a switch from the effector-independent cognitive learning phase to the effector-dependent motor learning phase after substantial practice.

We then situate internally-guided sequencing in a dual-process account of skill learning and propose computational analogues for the goal-directed and the habitual controller. We propose two hybrid reinforcement learning frameworks that integrate model-based and model-free mechanisms to account for the dual learning processes. Using simulations and model-fitting experiments, we compare the proposed hybrid frameworks, namely, value-of-information based arbitration and weighted-hybrid arbitration. We show that weighted-hybrid arbitration describes the empirical data better than other models. Our proposed framework gives a computational account of the learning in internally-guided sequencing.

Contents

List of Figures

List of Tables

Chapter 1

Introduction [1]

Our everyday experiences are an excellent demonstration of the surprisingly adaptive and fluid learning behavior that is orchestrated by the human brain. Such a learning behavior is a hallmark of human cognitive ability and spans a broad spectrum of tasks. Ranging from complex tasks such as cycling and driving to seemingly simpler ones such as typing and grasping movements, all tasks involve the acquisition of skillful behavior. Skill learning is a natural behavioral phenomenon concerned with the acquisition of the ability to perform tasks proficiently. *Motor skill learning* refers to learning a specific subclass of skills that involve sequential motor movements such that they are executed accurately and quickly with practice (Clegg et al., 1998; Haibach et al., 2018; Newell, 1991; Schmidt et al., 2019). Much of the early interest in motor sequencing focused on investigating the typical behavioral phenomenon in sequence learning tasks (Fitts and Posner, 1967; Hebb, 1961; Lashley, 1951). This has led to the formulation of many serial order canonical experimental tasks such as the $m \times n$ task (Bapi et al., 2000, 2006; Hikosaka et al., 1995) and discrete sequence production (DSP) task (Abrahamse et al., 2013; Verwey, 2001; Verwey et al., 2015) in the explicit domain and serial reaction time (SRT) task (Nissen and Bullemer, 1987; Robertson, 2007; Willingham, 1999) in the implicit domain. While explicit learning involves conscious awareness of what is being learned, implicit learning occurs without conscious awareness of learning. Subsequent research has extensively used these paradigms to understand the brain processes involved in sequence learning, memory, attention, etc.

In SRT and DSP tasks (see Figure 1.1), the participants repeatedly respond to a fixed set of visual stimuli organized in successive trials. Each trial involves presenting a sequence of visual cues that prompt corresponding keypress responses on a visuospatially-compatible button-box. In the $m \times n$ task (see Figure 1.2), each trial consists of n consecutive visual stimuli (called a hyperset). Each visual stimulus consists of m illuminated squares on a 3×3 grid presented on a screen. The participants learn to press m corresponding keys (called a set) successively in the correct order on a keypad in response to the visual stimulus. The visual stimuli that guide the sequencing behavior in such paradigms are predetermined and fixed by experimental design. The sequence of motor actions to be performed is not con-

[1] This chapter is a slightly modified version of our publication **Cognitive and Motor Learning in Internally-Guided Motor Skills**; Bera, K., Shukla, A., & Bapi, R. S. (2021) in *Frontiers in Psychology*.

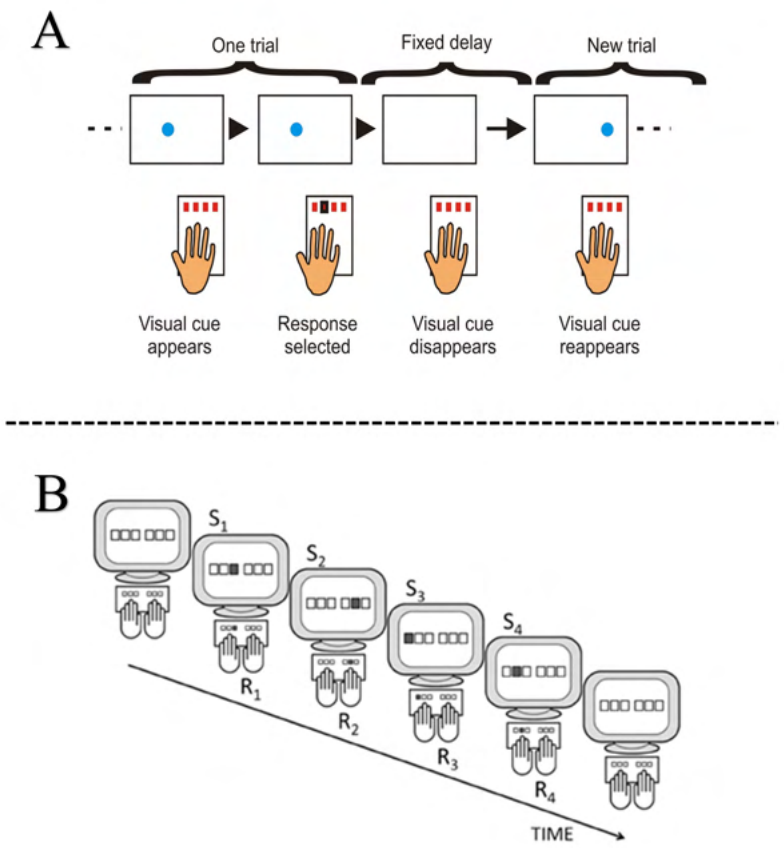

Figure 1.1 A schematic of the Serial Reaction Time (SRT) task (Fig. A; Robertson (2007)) and Discrete Sequence Production (DSP) task (Fig. B; Abrahamse et al. (2013)).

tingent on the participant's choice or plan. Therefore, these canonical tasks belong to a class of discrete sequence learning tasks that involve *externally-specified* or *externally-guided* (visual) sequences. The sequence of motor actions in such tasks is conditioned on fixed, externally-specified visual cues/stimuli.

While such simple canonical paradigms are useful for investigating skill learning in controlled experimental settings, they fail to account for a larger class of real-life motor tasks. Unlike SRT, DSP, or $m \times n$ task, many real-life motor skills are *internally-guided*, i.e., the sequence of the motor actions is triggered by self-choice or some internal model of the environment. Such tasks constitute a class of internally-guided motor tasks. The sequence of actions is self-initiated or generated internally by the participant and is not extrinsically prescribed or predetermined by the environment. Unlike externally-specified sequencing, the sequential action in such tasks is not elicited as a chain of stimulus-response pairs. While the visual cues might help the agent make sense of the environment in such tasks, it does not specify the sequence of motor movements to be executed. The central point of difference between externally-specified and internally-guided sequencing is that the latter involves volitional planning of motor action sequences. A template tracing task on paper is an example of an externally-specified task.

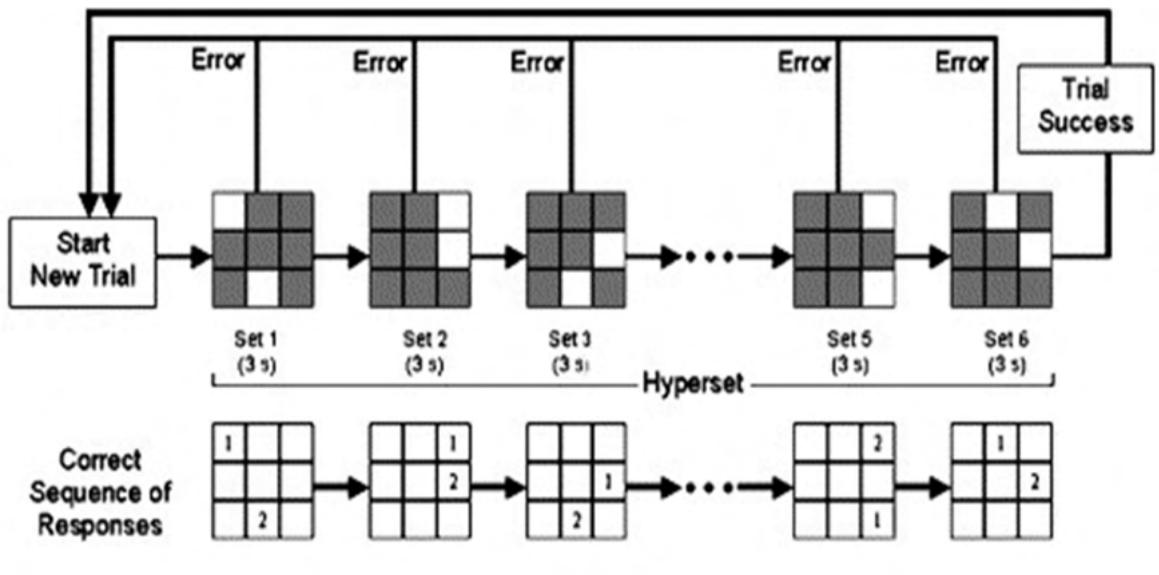

Figure 1.2 A schematic of the $m \times n$ task (Bapi et al. (2006)).

It employs external cues and visual feedback with a greater role of visuomotor associations for imitating the given template. On the other hand, drawing is an internally-guided task that relies on internal cues for guiding the pencil strokes to self-determined positions on paper. Such behavior is characterized by greater demands on brain processes related to memory and planning as compared to the tracing task. Other examples of such motor skills are composing music on a keyboard, creating a dance choreography, or solving a Rubik's cube. Such tasks involve planning as well as execution of a self-generated sequence of motor actions. The performance in internally-guided sequencing tasks depends on the dexterity of executing the motor actions and the ability to program the sequence of future actions.

Previous studies have investigated the motor behavior in externally-guided and internally-guided tasks and determined the neural underpinnings of the underlying processes. The externally-guided movements predominantly involve brain areas related to sensory guidance and optimization of movements, perception, and salience, whereas internally-guided movements involve brain areas related to muscle/movement selection, mental imagery, and planning complex behaviors (Drucker et al., 2019; Gowen and Miall, 2007). Other investigations have confirmed the role of cerebellar and premotor circuits in externally-guided tasks and basal ganglia, pre-supplementary motor cortex and dorsolateral prefrontal cortex in internally-guided tasks (Jueptner et al., 1996; Jueptner, 1998; van Donkelaar et al., 1999).

In externally-specified sequencing, bindings between the presented stimuli and the corresponding responses emerge with simple association rules between stimuli and response (S-R). Selecting an action in response to a given stimulus binds the codes of the action-relevant stimulus attributes and the corresponding action codes (Logan, 1988). Due to repeated execution of sequences, the activity of the system controlling stimulus-based actions results in stimulus-response or sensorimotor learning (Her-

	Serial reaction time task (SRT)	Discrete sequence production task (DSP)	*m* × *n* task	Grid-sailing task (GST)
Features of experimental paradigm				
Number of effectors	1/2	2	1	1
Number of choices/fingers used	4	4/6/8	3	3
Stimuli	Visual: spatially-compatible and key-specific	Visual: spatially-compatible and key-specific	Visual: consists of *m* illuminated squares on a grid	Visual cues for start, goal and agent positions
Sequence length	10	3–8	10–12	5–7
Number of trials	800	500–1,000	10–20 successful trials	20
Behavioral measures	Response time	Response time	Choice time, movement time	Reward, number of moves, execution time, reaction time
Nature of sequences and learning				
Sequence specification	Explicitly specified	Explicitly specified	Explicitly specified—discovery by trial and error	Internally planned
Kind of sequences learnt	Typically first-order and second-order	Typically first-order and second-order	Hierarchical sequence	Higher-order trajectory of grid-cell states
Nature of learning	Implicit	Explicit	Explicit	Explicit

Figure 1.3 Task comparison between externally-specified (SRT, DSP, $m \times n$) and internally-guided sequencing tasks.

wig and Waszak, 2009). Therefore, the sequencing in the externally-specified domain is exhibited as a chain of stimulus-response-effect (S-R-E). On the other hand, the internally-guided or voluntary actions typically involve a goal-directed motivation to achieve an internally pre-specified outcome. The studies have shown that such self-determined action goals play a role in the acquisition and planning of internally-guided actions (Hommel, 2003; Hommel et al., 2001). The activity of the system guiding intention-based actions results in action-effect or ideomotor learning due to the formation of associations between movements and their ensuing sensory effects (Herwig and Waszak, 2009). According to the ideomotor framework of action control (Greenwald, 1970; Prinz, 1997), internally-guided actions primarily refer to anticipated action effects or, in other words, response-stimulus (R-S) bindings. In internally-guided actions, the participants might only attend to response-effect (R-E) contingencies (Herwig and Waszak, 2009). In light of these differences, none of the previous studies have explored the nature of learning processes in such a class of discrete, self-guided sequential movement tasks. Motivated by this apparent gap, this work investigates skill learning in internally-guided sequencing.

Sequence learning in simple grid-navigation tasks is an example of an internally-guided sequencing paradigm. The tasks involve navigating (typically, using a cursor) on the grid from the *start* position to the *goal* position. Each unique trajectory from the start to the goal position constitutes a novel sequence of keypresses. The optimality of trajectory is conditioned on the task specifications such as the reward scheme, possible agent movements, and time constraints. Participants are free to choose among many possible optimal trajectories for a trial to be successful. The repeated execution of these trajectories results in learning a self-generated, voluntary sequence of keypresses. The behaviors in grid-navigation tasks give us rich insights into the learning processes involved in internally-guided sequencing. We propose a novel usage of the simple grid-navigation task - *Grid-sailing task* (GST; Fermin et al. (2010,

2016)) as a canonical paradigm to investigate the learning processes involved in internally-guided sequencing. The GST requires navigating a 5 × 5 grid from start to goal position (referred to as the SG position) using a given key-mapping (KM). The KM associates possible movement directions of the cursor with the corresponding keyboard buttons. The participants are instructed to reach the goal in an optimal number of steps as quickly as possible. Figure 1.3 provides a concise summary of different sequencing tasks—SRT, DSP, $m \times n$, and GST—based on the experimental paradigm and the nature of learning involved.

Using GST as our canonical paradigm, we present an empirical and computational investigation of internally-guided sequencing. The most of the practical and everyday motor skills are internally-guided and therefore, the significance of this work lies in the fact that it tries to systematically understand the cognitive mechanisms underlying internally-guided skill learning.

Outline of the thesis: This thesis is organized into two sections. The first section presents empirical investigation of the behavioral phenomenon in internally-guided sequencing. The second section presents a computational account of learning in GST.

- In the first section, chapter 2 investigates the different learning processes involved in internally-guided sequencing. The participants performed GST on two conditions - Single-SG and Mixed-SG. We analyzed the sequence-specific performance improvements to test for motor learning in the Single-SG condition. We further analyzed performance improvements due to the transfer of KM-specific internal model in the Mixed-SG condition to show evidence for the role of cognitive learning in internally-guided sequencing.

- Chapter 3 investigates the nature of motor learning in GST. Motor learning can can result from two processes - motor adaptation or motor chunking. The study provides evidence for practice-driven performance improvements in GST due to motor chunking.

- Chapter 4 probes inter-manual transfer of skills. The participants performed a transfer-task over three sessions on different days. We analyzed the performance on normal and transfer blocks to show evidence for the cognitive to motor 'switch'.

- In the second section, chapter 5 describes a reinforcement-learning based computational account of learning in GST. We describe a computational equivalence between the cognitive–motor and the model-based–model-free dichotomies. We further show simulations and compare the model fits of various reinforcement learning algorithms on the experimental data from chapter 2.

- The thesis ends by summarizing the main findings, highlighting its significance and outlining the future directions.

Chapter 2

Cognitive and Motor Learning in Internally-guided Sequencing[1]

2.1 Introduction

We considered the involvement of two learning components—motor and cognitive. The cognitive component involves learning the sequential order of movements, whereas the motor component concerns the acquisition of fine-tuned movement dynamics and sensorimotor integration (Doya, 2000; Ghilardi et al., 2009; Penhune and Steele, 2012). Using GST as our canonical paradigm, we employ two behavioral experiments to identify the underlying learning processes in the internally-guided sequencing. In Experiment-1 (Single-SG condition), participants perform GST on a single SG-condition. We show evidence for motor learning due to the repeated execution of sequences. In Experiment-2, the participants use the learned KM from Experiment-1 to perform grid-navigation on the Mixed-SG condition, which consists of randomized trial order of two previously unseen SG conditions. A successful transfer of a KM-specific internal model would enable efficient trajectory planning on the novel SG conditions and, thus, would point out the role of the cognitive learning in Experiment-2. We further make a case for using GST-like grid-navigation tasks for investigating the typical behavioral phenomena in internally-guided sequencing.

2.2 Experiment-1: Single-SG Condition

We hypothesize that sequence-specific motor learning contributes to the learning in GST. As the participants repeatedly execute the same trajectory, the motor movements are optimized to facilitate accurate and fast sequential keypresses. This can be empirically tested by examining the effect of trials on the mean execution in Experiment-1 (also referred to as the Single-SG condition). The Single-SG condition also involved a rotation trial to test whether the learning in GST occurs due to the acquisition of a motor program or general motor improvements. The general motor improvements can result from factors such as task familiarity or adaptation. The rotation was introduced such that the sequence of

[1]This chapter is a slightly modified version of our publication **Cognitive and Motor Learning in Internally-Guided Motor Skills**; Bera, K., Shukla, A., & Bapi, R. S. (2021) in *Frontiers in Psychology*.

keypresses required to navigate the cursor from the start position to the goal position remained the same as in the normal trials. Consequently, the execution time on the rotation trials is expected to remain unaffected if the performance improvements in GST occur only due to general motor improvements.

2.2.1 Methods

2.2.1.1 Participants

Forty-two healthy participants volunteered for the study. The participant pool consisted of 29 women and 13 men between ages 17 and 27 years (mean = 21.02; SD = 2.46) years. All participants were non-musicians with normal or corrected-to-normal vision. The study was approved by the Institute Review Board, IIIT-Hyderabad, India. The participants gave informed written consent before the study. Additionally, permission for participation was obtained from the College Principal for participants below 18 years of age. The participants initially performed Experiment-1 (Single-SG condition with visuomotor rotation trial) followed by Experiment-2 (Mixed-SG Condition).

2.2.1.2 Apparatus

The participants were seated on a chair facing a high-resolution 24-in computer screen placed approximately 2 ft away. A conventional desk keyboard was used to record responses. The participants used the right index, middle, and ring fingers to press the number-pad buttons "4," "5," and "6," respectively. All the other keys on the number-pad were removed to prevent meddling in response selection. The experiment program for stimulus presentation and data recording was written using Python3 and PyGame (Python Game Development[2]).

2.2.1.3 Procedure

The participants were verbally instructed about the task procedure before the session started. A 5×5 grid with a red fixation cross at the center was displayed at the beginning of each trial. On pressing the "space" button, after a random delay of 500–1,000 ms, the trial started with the start position marked as a green tile and the goal position marked as a blue tile. The cursor, shown as a black triangle, was initially placed in the starting position. The participants were given 6 s to solve each trial, and this duration was not explicitly conveyed to them. During the trial response period, participants executed sequential keypresses to navigate the cursor from the starting position to the goal position. The possible cursor-movement directions were defined by the KM (see Figure 2.1 A). In the beginning, the task required participants to explore the KM directions and its association with the corresponding keys by trial and error.

The participants were explicitly instructed to achieve a maximum score (of 100 points) while executing each trial as quickly as possible. If an optimal path is traversed, a maximum of 100 points

[2]Retrieved from https://www.pygame.org

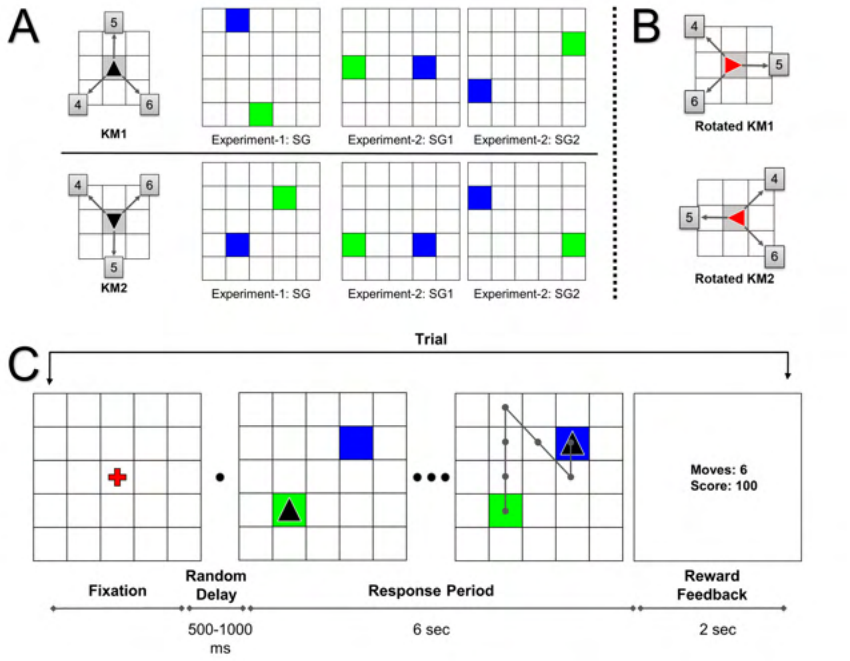

Figure 2.1 (A) Key-mapping (KM) and start-goal (SG) position sets used in the experiment. Each participant was randomly assigned either KM1 or KM2. The boxed numbers on KM figure show corresponding numeric keys associated with the movements. In SG figures, green and blue tiles represent start and goal positions, respectively. (B) The 90° clockwise rotated KMs used in the rotation trials in Experiment-1. (C) Task diagram: sequence of trial events (adapted from Fermin et al. (2010)). In this illustration, the participant is assigned key-map KM1. An example optimal trajectory is shown on the grid.

is awarded for that trial. A minimum steps trajectory from start to the goal is considered an optimal trajectory. If the participant took a non-optimal path, a penalty of -5 points incurred for every excess move. In case the participant tried to perform an invalid move, such as moving out of the grid, the cursor position remained the same with an incremented move count. If the participant failed to reach the goal in the given time duration, 0 points were awarded for that trial. At the end of each trial, the performance feedback was displayed for 2 s, following which the fixation screen signaled the beginning of the next trial. On the center of the feedback screen, the performance feedback was presented as two numbers. The display showed the number of moves in the traversed trajectory and the reward score for that trial. A trial outline is shown in Figure 2.1 C. The participants were given a rest block after every 20 trials to minimize the effects of muscle fatigue on the performance. The participants were also advised to maximally re-use the explored trajectories in order to execute the task quickly and accurately.

Two different KMs were used in the experiment to avoid any unwanted performance effects or bias due to a particular KM. Moreover, each KM was associated with a unique set of SG pairs (see Figure 2.1 A). The participants were randomly assigned one of the two possible KMs. The participants used the

same assigned KM throughout the experiment for both, Single-SG and Mixed-SG conditions. Twenty-four participants used KM1, whereas eighteen participants used KM2 for the experiment.

In Experiment-1, the participants repeatedly performed GST on a single SG condition. The participants were presented with the same SG condition for trials 1–41. The rotation trial (trial 42) was introduced after the completion of 41 successful trials. The rotation trial was followed by the re-introduction of the learned single SG condition for the next five trials (trials 43–47). The post-rotation trials (trials 43–47) were used only for a comparative analysis between the rotation and normal condition. The rotation trial involved a 90° clockwise rotation of the grid. The start and goal positions also changed accordingly with the grid rotation. The rotated cursor changed its color from black to red to indicate the transformed KM associations (see Figure 2.1 B). Therefore, the sequence of keypresses required to reach the goal position effectively remained the same. In the case of error trials in the rotation condition, the participants were repeatedly presented with the rotation trial. The participants were already instructed about the rotation trial beforehand. The participants took about 15 min to complete Experiment-1.

2.2.1.4 Behavioral Measures

The number of moves in the traversed trajectory, reward obtained, reaction time and execution time were the performance measures recorded for each trial of the experiment. Reaction time is defined as the time interval between the onset of stimuli and the first keypress. Execution time is the total time taken for sequential keypresses in a particular trial. Execution time is computed as the difference between the keypress time of the last and the first response. For analysis purposes, the trials were classified into three categories (1) Successful trials—if the goal position is reached with a non-zero reward, (2) Optimally successful trials—if the goal position is reached in an optimal number of moves and thereby scoring a maximum reward, and (3) Error trials—if the goal position is not reached in the given time duration.

2.2.2 Results

The following behavioral measures were included for the analysis: reward score, reaction time, execution time, number of moves, and error rate. The error rate is a computed measure that denotes the average number of error trials attempted to complete one successful trial. The successful and error trials were both included in the analysis to show the emergence of learning and skillful behavior in the task. However, only successful trials were considered for other analysis purposes. A within-subjects repeated-measures ANOVA was used to test for the effect of practice (trials) on different behavioral measures. A series of Wilcoxon signed-rank tests was performed on various measures to compare the performance on rotation and normal trials. Repeated-measures ANOVA was used to probe any KM-specific effects on the performance. The statistical analysis was performed using Python (scipy and statsmodels packages) and JASP software (JASP Team, 2020).

The learning in the task is evident from the performance improvements in various behavioral measures. With practice, we see an increasing and decreasing trend in reward and execution time, respec-

tively, which suggests that within a few (10–15) trials, the participants progressively learned to perform the task while optimizing for speed (execution time) and accuracy of navigation (reward; see Figure 2.2 A). We took reward, moves, execution time, and reaction time as dependent measures of learning for successful trials. To evaluate the learning behavior, we plotted the mean values of behavioral measures in successful trials (see Figures 2.2 B,C). The mean reward increases to a maximum of 100 points as the number of moves reduces over the practice to reach the optimal/minimum number of steps. A non-parametric Friedman test of differences among repeated measures (within-subjects) rendered a significant effect of trials on average reward obtained $(\chi^2(40) = 73.97, p < 0.001)$ and the average number of moves required to reach the goal $(\chi^2(40) = 73.97, p < 0.001)$.

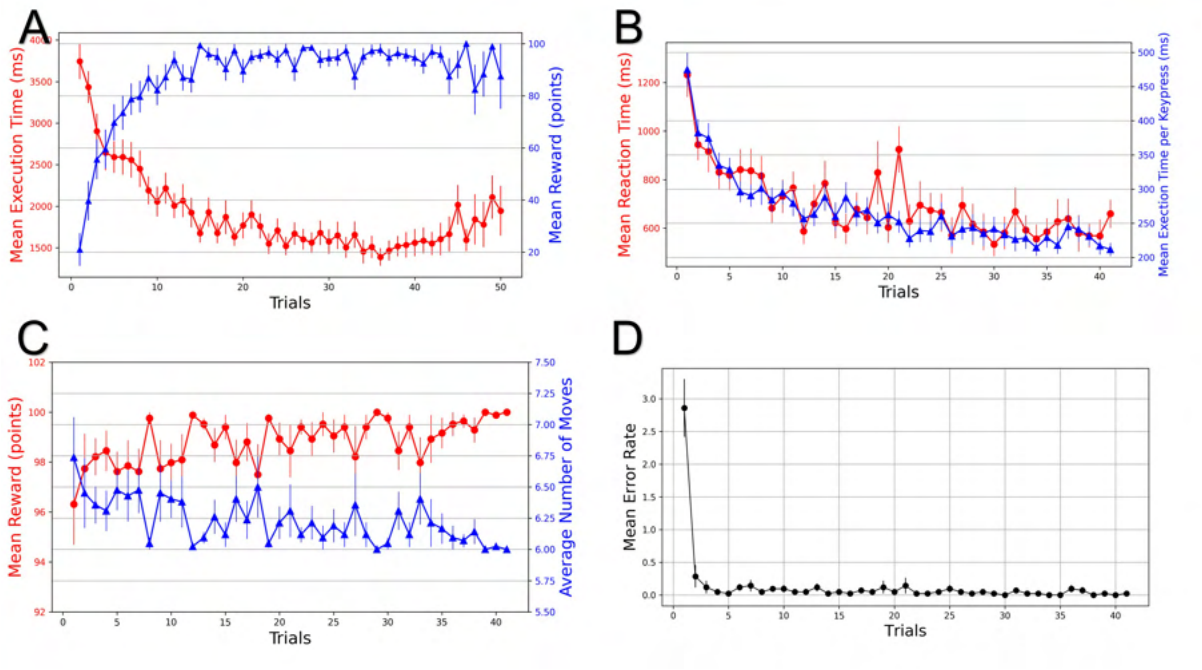

Figure 2.2 Trial-by-trial course of performance improvement in Single-SG condition (rotation trial excluded). The bars on the plot data-points denote standard error in measurement. (A) Evolution of learning behavior in the task. Mean execution time and mean reward across trials—averaged over both successful and error trials. (B) Mean reaction time and normalized execution time in successful trials. (C) Mean reward and average number of moves in successful trials. (D) Mean error rates in successful trials.

The learning is also evident by comparing the mean execution time of the first successful trial (M = 3,110 ms, SD = 1,062) with the last successful trial (M = 1,271 ms, SD = 380 ms). The mean reaction time decreased from 1,232 ms (SD = 592) to 659 ms (SD = 380). The Friedman test indicated significant improvements in execution time $(\chi^2(40) = 485.90, p < 0.001)$ as well as reaction time $(\chi^2(40) = 300.55, p < 0.001)$. However, this decrease in execution time could have been a function of the number of moves in the trajectory. Therefore, we computed normalized execution times or execution time per keypress to account for the unequal lengths of trajectories in successful trials. The Friedman

test indicated significant improvements in normalized execution time ($\chi^2(40) = 499.93, p < 0.001$). The acquisition of learned sequences was examined by computing the error rates and plotting them against trials. A steep decrease in error rates is observed over the first few trials (see Figure 2.2 D). Additionally, sequence-specific motor learning was examined by controlling the number of keypresses and the trajectories followed. For each participant, the most frequently used optimal trajectory was determined. The trials that employed the most-frequented optimal trajectory were extracted. A decrease in mean execution time from 2,473 ms (SD = 1,051) to 857 ms (SD = 157) in extracted trials suggests sequence-specific learning. To evaluate the performance improvements across these trials, we performed a Friedman test, which indicated a significant effect of trials ($\chi^2(38) = 78.72, p < 0.001$) on execution time.

In order to examine whether the learning observed in the GST was particular to KM, we performed 2 (KM: 1 and 2) × 41 (Trials: 1–41) mixed repeated-measures analysis of variance (ANOVA) on normalized execution time. The KM was used as a between-subject factor, and trials were a within-subject factor. A Greenhouse-Geisser correction was applied when the ANOVA assumptions were violated. The ANOVA results suggested a significant main effect of trials ($F(11.16, 446.25) = 19.148, p < 0.001, \eta_p^2 = 0.324$) on the normalized execution times. Similarly, a significant main effect of KM ($F(1, 40) = 7.517, p = 0.009, \eta_p^2 = 0.158$) indicated that the normalized execution times are different for the two KM. However, the Trial × KM interaction was not found to be significant ($F(11.16, 446.25) = 1.347, p = 0.194, \eta_p^2 = 0.033$), suggesting that the variation in normalized execution time across the trials is not dependent on KM.

On the visuomotor rotation trial (trial 42), we observed a spike in the execution time (see Figure 2.3 A). The execution time comes down with the re-introduction of the learned SG condition after the rotation trial. To assess whether the mean execution time for the visuomotor rotation trial is significantly higher than the normal condition, we took the average execution time of the preceding and the succeeding optimally successful trials and compared it with the rotation trial (trial 42). The mean execution time increased from 1,473 ms (SD = 496) in the normal trials to 2,906 ms (SD = 919) in the rotation trial. A Wilcoxon signed-rank test was used as the normality assumptions were violated. It suggested that the mean execution time for the rotation trials is significantly higher than the normal trials ($df = 29, Z = 0, p < 0.001$). Similarly, the mean reaction time increased from 900 ms (SD = 420) in the normal trials to 1,176 ms (SD = 672) in the rotation trial (see Figure 2.3 B). The Wilcoxon signed-rank test indicated that the difference in mean reaction time on the rotation trial and normal trials was significant ($df = 41, Z = 245.50, p = 0.010$). On following a similar procedure, we found that the mean reward obtained decreased from 99.702 (SD = 1.581) to 95.595 (SD = 9.513) on the rotation trials (see Figure 2.3 C). The Wilcoxon signed-rank test suggested that the difference in mean reward score obtained on normal and rotation trials is significant ($df = 41, Z = 96.50, p = 0.006$). Similarly, the number of moves executed increased from 6.060 (SD = 0.316) in the normal trials to 6.881 (SD = 1.903) in the rotation trials (see Figure 2.3 D). The Wilcoxon signed-rank test suggested that the increase in the number of moves is significant ($df = 41, Z = 8.50, p = 0.006$). The error rates also increased

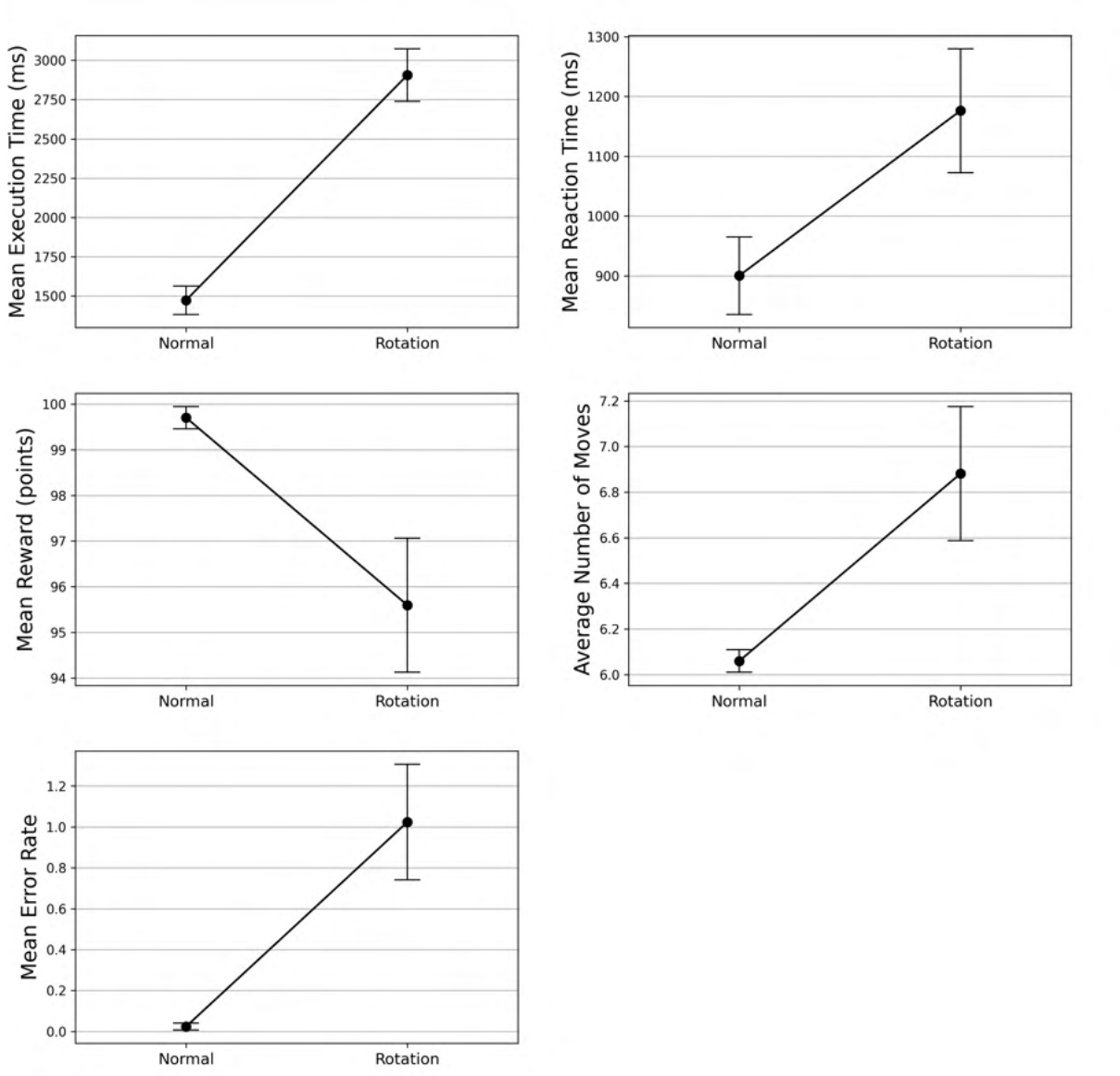

Figure 2.3 Comparison of performance on normal and visuomotor rotation trials in Experiment-1. The bars on the plot data-points denote standard error in measurement. (A) Mean execution time on optimally-successful rotation trials is significantly higher than the average of preceding and succeeding optimally-successful trials. (B) Mean reaction time on successful rotation trials is significantly higher than the average of preceding and succeeding successful trials. On the rotation trial, the average reward obtained (C) is significantly lesser while the average number of moves (D) is significantly higher. (E) The mean error rate also increases in the rotation trials.

from 0.024 (SD = 0.108) to 1.024 (SD = 1.828) in the rotation trials (see Figure 2.3 E). The Wilcoxon signed-rank test also suggested a significant difference in error rates ($df = 41, Z = 0, p < 0.001$) in both conditions.

2.2.3 Discussion

In line with previous GST studies and other skill learning tasks, the performance improvements in terms of speed (execution time, reaction time) and accuracy (reward) suggest the acquisition of skillful behavior (Abrahamse et al., 2013; Fermin et al., 2010; Hikosaka et al., 1995; Nissen and Bullemer, 1987; Sakai et al., 2003; Willingham, 1999). The practice-driven performance improvements in various behavioral measures in Single-SG trials suggest overall learning in GST. The practice-driven performance improvements in execution time provide evidence for motor learning that occurs due to the fine-tuning of motor movements. The Single-SG condition also included a visuomotor rotation trial to probe if the performance improvements in GST can be solely attributed to general motor improvements. The performance degradation on execution time and other measures such as reward, reaction time, and error rate suggests that the performance improvements in GST can be attributed to the sequence-specific learning processes (specifically, the acquisition of the motor program).

2.3 Experiment-2: Mixed-SG Condition

In Experiment-1, the participants repeatedly performed GST on a single SG condition using the same KM. The participants not only learned the motor program associated with the sequence of movements to reach the goal position but also internalized the navigation strategies related to the specific KM. The results of Experiment-1 established trajectory-specific motor learning. Further, to investigate KM-specific cognitive learning, we designed Experiment-2 (also referred to as the Mixed-SG condition) to test the transfer of KM-specific learning to novel SG conditions. We anticipate that the transfer of KM-specific learning will lead to efficient trajectory planning on novel SG conditions. The account can be empirically tested by comparing various performance measures during the initial trials of Experiment-1 and Experiment-2.

2.3.1 Methods

2.3.1.1 Participants

All the participants performed Experiment-2 after completing Experiment-1.

2.3.1.2 Apparatus

The experimental setup and apparatus were the same as in Experiment-1.

2.3.1.3 Procedure

In the Mixed-SG condition, the general task paradigm was the same as in Experiment-1 except for the SG-conditions. In the Mixed-SG condition, participants employed the previously learned KM (from Experiment-1) to perform grid-navigation on two novel SG conditions. The optimal number of steps in both the SG conditions were the same. During the experiment, each trial was randomly assigned to one of the two possible SG conditions. The participants performed GST on the randomized and mixed order of SG conditions. The experiment terminated when the participant performed 20 successful trials of each SG condition. The participants took about 15 min to complete the Mixed-SG condition task.

2.3.1.4 Behavioral Measures

The behavioral measures logged and analyzed were the same as in Experiment-1.

2.3.2 Results

A within-subjects repeated-measures ANOVA was used to test for the effect of practice (trials) on different behavioral measures. A series of Wilcoxon signed-rank tests was performed on various measures to test for the transfer of learning in Experiment-2. Repeated-measures ANOVA was used to probe any KM-specific or SG-specific effects on the performance. The first 20 successful trials of each SG condition were considered for analysis. Mean execution times and reaction times for a total of 40 successful trials were plotted against the trials. We observe that with practice, the participants become more accurate and efficient in performing GST on the Mixed-SG condition (see Figure 2.4 A). Both the execution time and reaction time, as dependent measures of performance, decrease with practice (see Figure 2.4 B). The learning is evident by the decrease in the mean execution time from 2,854 ms (SD = 971) in the first successful trial to 1,298 ms (SD = 428) in the last successful trial. And the mean reaction time decreased from 1,278 ms (SD = 623) to 733 ms (SD = 403). To evaluate whether the change across the trials is statistically different, we performed a non-parametric Friedman test of differences among repeated measures (within-subjects) for trials 1 through 40. We observed a significant effect of trials on the mean execution time ($\chi^2(39) = 423.35, p < 0.001$) as well as the mean reaction time ($\chi^2(39) = 210.40, p < 0.001$). A Friedman test also indicated a significant effect of trials ($\chi^2(39) = 469.06, p < 0.001$) on normalized execution time. The mean reward scores improved from 97.50 (SD = 5.325) in the first trial to 99.52 (SD = 1.851) in the last trial. The effect of trials was significant on the mean reward obtained ($\chi^2(39) = 61.96, p = 0.011$; see Figure 2.4 C). The average number of moves required to reach the goal position decreased from 6.50 (SD = 1.065) to 6.095 (SD = 0.370) with practice. A Friedman test indicated a significant effect of trials on the average number of moves ($\chi^2(39) = 61.96, p = 0.011$; see Figure 2.4 C). A Friedman test on mean execution time in optimally successful trials rendered a significant effect ($\chi^2(39) = 191.42, p < 0.001$) of trials. The mean error rates were computed by averaging the participant error rates while preserving the trial order. A steady decrease in error rates is observed with practice (see Figure 2.4 D).

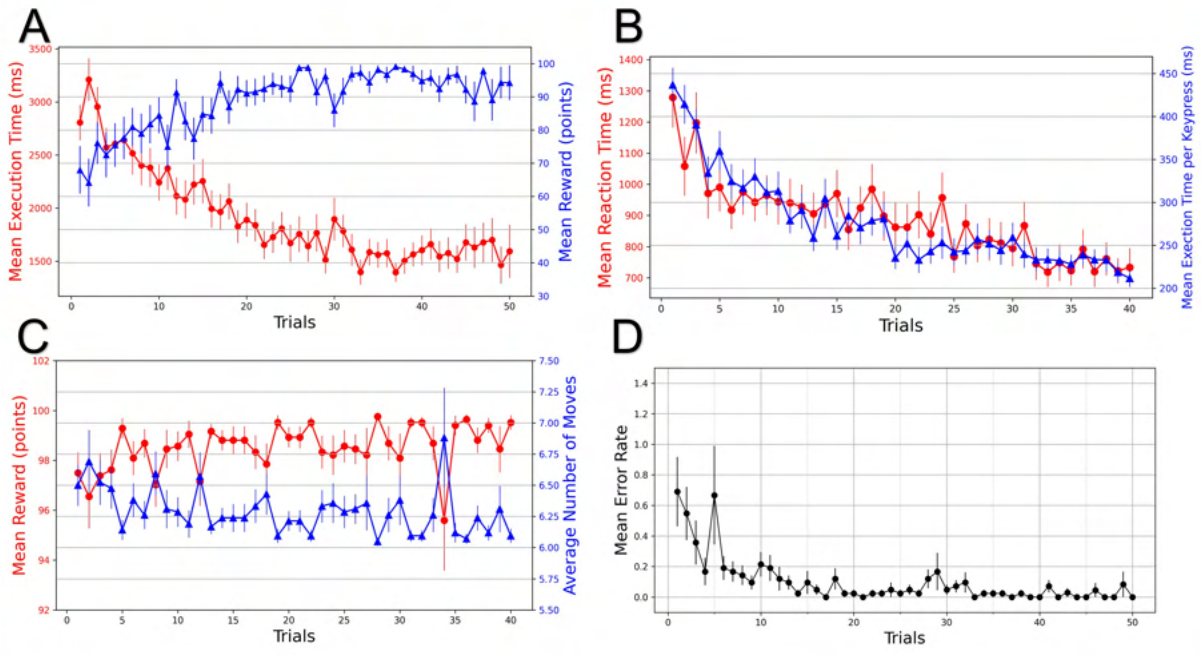

Figure 2.4 Trial-by-trial course of performance improvement in Mixed-SG condition. The bars on the plot data-points denote standard error in measurement. (A) Evolution of learning behavior in the task. Mean execution time and mean reward across trials—averaged over both successful and error trials. (B) Mean reaction time and normalized execution time in successful trials. (C) Mean reward and average number of moves in successful trials. (D) Mean error rates in successful trials.

Additionally, we examined if the performance in the Mixed-SG condition was particular to KM. We performed 2 (KM: 1 and 2) × 40 (Trials: 1–40) mixed repeated-measures analysis of variance (ANOVA) on normalized execution time with KM as a between-subject factor and the trials as a within-subject factor. A Greenhouse-Geisser correction was applied when the ANOVA assumptions were violated. The ANOVA results suggested a significant main effect of trials ($F(13.09, 523.48) = 17.326, p < 0.001, \eta_p^2 = 0.302$), indicating that the normalized execution time varies across the trials. A non-significant main effect of KM ($F(1, 40) = 0.782, p = 0.382, \eta_p^2 = 0.019$) indicated that the normalized execution times are not different for the two KM. Moreover, a non-significant Trial × KM ($F(13.09, 523.48) = 1.325, p = 0.193, \eta_p^2 = 0.032$) interaction suggested that the variation in normalized execution time across the trials is not dependent on KM.

Since the participants used the same KM assignment as that in the Single-SG condition, we anticipate the transfer of learning to occur from the Single-SG condition to the Mixed-SG condition. We analyzed behavioral measures such as error rate, reward, and execution time for the first trial in both conditions to probe for transfer effects. The mean error rate improved from 2.857 (SD = 2.859) in the Single-SG condition to 0.690 (SD = 1.473) in the Mixed-SG condition. A Wilcoxon signed-rank test comparing the error rates during the first trials of both conditions reported significant differences ($df = 41, Z = 508.00, p < 0.001$). The mean reward score improved from 20.952 (SD = 40.714) in the Single-SG

condition to 67.976 (SD = 46.224) in the Mixed-SG condition. A Wilcoxon signed-rank test revealed that the mean reward obtained is significantly higher ($df = 41, Z = 36.00, p < 0.001$) for the initial trial in the Mixed-SG condition as compared to the first trial in the Single-SG condition. Similarly, the mean execution time improved from 3,743 (SD = 1,380) ms to 2,807 (SD = 1,083) ms due to the transfer effects. A Wilcoxon signed-rank test suggested that the mean execution time for the first trial in the Single-SG condition is significantly different ($df = 41, Z = 743.00, p < 0.001$) as compared to the first trial of the Mixed-SG condition.

The Single-SG condition does not require participants to employ all three keys to reach the goal position. In both the KM groups, the participants only needed keys 4 and 5 to build the trajectory from the start to the goal position. Therefore, in both the KM groups, at the end of the Single-SG condition, the participants are highly trained with the response effects for two (keys 4 and 5) of the three keys. In the Mixed-SG condition, condition SG1 requires using keys 5 and 6 to build an optimal trajectory, whereas condition SG2 requires keys 4 and 5 to navigate to the goal position. Since the participants are highly trained on response-effect contingencies for keys 4 and 5 but not for key 6, the differential amount of practice may benefit performance on condition SG2 (employing keys 4 and 5) but not condition SG1 (employing keys 5 and 6). Thus, one can argue that performance will be influenced by a differential amount of practice based on SG conditions in the Mixed-SG condition. To probe this, we analyzed the effect of SG on execution time in the Mixed-SG condition. We performed 2 (SG: 1 and 2) × 20 (successful trials: 1–20) repeated-measures ANOVA on the mean execution time for each KM. A Greenhouse-Geisser correction was applied when the ANOVA assumptions were violated. For KM1, the ANOVA results reported a main effect of trials ($F(4.875, 112.123) = 27.583, p < 0.001, \eta_p^2 = 0.402$), suggesting practice-driven learning. The test reported a non-significant main effect of SG ($F(1, 23) = 3.915, p = 0.060, \eta_p^2 = 0.006$) and Trial × SG interaction ($F(6.983, 160.601) = 1.081, p = 0.378, \eta_p^2 = 0.010$). Similarly, for KM2, the ANOVA results reported the main effect of trials ($F(19, 323) = 13.192, p < 0.001, \eta_p^2 = 0.263$), suggesting practice-driven performance improvements. It suggested a non-significant main effect of SG ($F(1, 17) = 0.035, p = 0.854, \eta_p^2 = 0.0001$) and Trial × SG interaction ($F(19, 323) = 1.204, p = 0.252, \eta_p^2 = 0.023$). The results suggest that the execution time is not different for the two SGs in the Mixed-SG condition in both the KM groups. Therefore, the performance is not influenced by the differential amount of practice based on SG conditions in the Mixed-SG condition.

2.3.3 Discussion

The randomized and mixed order of SG conditions in Experiment-2 minimized the trajectory-specific performance improvements that occur due to the repeated execution of the keypress sequences. Significant performance improvements were observed for normalized execution time in successful trials and execution time in optimally-successful trials. Other performance measures such as reward and reaction time also improved with practice. This efficient performance of GST on new and randomly-ordered SG conditions can be attributed to the ability to use a previously learned KM-specific internal model

for planning navigation strategies. As the learned KM relations can be successfully applied to new SG conditions, the participants could generalize the learning from the Single-SG condition to the Mixed-SG condition. The positive transfer effects support the idea of cognitive learning because the participants cannot simply transfer a learned motor sequence in the Mixed-SG condition.

In both the KM groups, the Single-SG condition employed only two (keys 4 and 5) of the three possible cursor movements to construct an optimal trajectory. One can raise the question that the transfer of learning would be better on the novel SG condition that employed the same practiced keys (namely, condition SG2) compared to the other novel SG condition (namely, condition SG1) in Experiment-2. However, the analysis revealed no difference in performance in execution time in both the SG conditions. This suggests that the transfer of the internal model related to the KM is not contingent on the specific keys that are practiced in various SG conditions.

2.4 General Discussion

We investigated the nature of learning in internally-guided sequencing. We argued that GST-like grid-navigation tasks are exemplars of such a paradigm and hypothesized the role of motor and cognitive learning processes in learning in GST. We proposed a novel use of GST in two behavioral experiments to this end. In Experiment-1 (Single-SG condition), we investigated the progressive nature of learning, as evidenced by improvements in various behavioral measures. We provide evidence for the role of trajectory-specific motor learning in GST by showing the effect of trials on execution time in the Single-SG condition. The performance degradation on the introduction of a visuomotor rotation trial suggests that the learning in GST involves the acquisition of a motor program and therefore, it cannot be solely attributed to the general motor improvements. In Experiment-2 (Mixed-SG condition), we provide evidence for the role of KM-specific, trajectory-independent learning in GST. The transfer-related performance improvements in the Mixed-SG condition provide evidence for the acquisition of a KM-specific internal model that translates as cognitive learning in GST.

2.4.1 General Stages of Learning in GST

Improvements in various behavioral measurements such as execution time, reaction time, and reward score indicate learning in GST (see Figure 2.2). In the Single-SG condition, the participants initially tried to learn the possible movement directions and the corresponding key-map (KM) by trial and error. As the participants became familiar with the association between keypresses and corresponding cursor movements, they learn the effects of their responses. In further attempts, using the learned KM, the participants execute the keypresses to move the cursor in the direction of the target. Further practice enables them to plan simple and optimal navigation strategies to reach the goal. In the late phase, the repeated execution of the optimal trajectory drives performance improvements due to motor learning. We anticipate the role of motor chunking, due to which the planned trajectory to the goal position

is segmented into sub-sequences of individual motor actions (keypresses). The late stage of practice would be characterized by an "automatic" mode of execution with reduced cognitive and attentional demands. Motor chunking would enable the participants to perform the sequence as a whole without relying on individual response-effect contingencies. The practice-driven performance improvements in GST due to chunking are investigated in another study (Bera et al., 2021b).

The trajectory planning was guided by feedback from the reward score and the number of moves (see Figure 2.2 A). The reward feedback gives a measure of the optimality of the trajectory followed. A steep decrease in the number of error trials (error rates) is observed after the first successful trial (see Figure 2.2 D). Further practice enabled planning of optimal trajectories, as evident from the increase in mean reward. The participants quickly hit the reward ceiling (reward = 100) within 10–15 trials, implying that they have learned to navigate optimally. After a substantial amount of practice, as the KM model and SG trajectories are thoroughly learned, the (reward) feedback became less consequential for task accuracy (see Figure 2.2 C). Nevertheless, we saw a further performance improvement in normalized execution time (task speed) of successful trials (see Figure 2.2 B). To control the number of moves over which the execution time is computed in successful trials, we performed the normalized execution time analysis. A statistically significant improvement in normalized execution time shows growing expertise in performing the sequence. While multiple optimal trajectories are possible for a given SG position, the performance improvements due to repeated execution of the same trajectory can be attributed to motor learning. Therefore, we also performed the execution time analysis with a control on the number of moves and the sequential keypresses of the trajectory traversed. A statistically significant improvement in execution time confirms the role of motor learning due to repeated execution of the same trajectories.

2.4.2 Cognitive Aspects of Internally-Guided Sequencing

In addition to motor learning, the performance in internally-guided paradigms is also contingent on the ability to plan the sequence of actions efficiently. In GST, determining the sequence of keypress execution corresponds to planning a trajectory from the start to the goal position. Such planning and trajectory-generation are analogous to the goal-directed behavior in the knight's tour on a chessboard. To reach a given goal position on the chessboard, goal-directed planning is employed to generate an optimal sequence of moves (analogous to a trajectory in GST) using an internal map based on possible movement directions of a knight (similar to a KM in GST). In both cases, the conceived reach pattern used for planning trajectories is KM-specific. The acquisition of such a KM-specific internal model helps in planning trajectories and amounts to cognitive learning in GST.

Therefore, we hypothesized that cognitive learning processes contribute to the learning in GST. The role of cognitive learning could be confirmed if the participants can generalize the learning from a learned SG condition to other novel SG conditions. To test this, we performed Experiment-2, where we asked the participants to perform GST on randomized and mixed order of novel SG positions using the learned KM from Experiment-1. The transfer-related performance improvements in various behavioral measures confirm the role of cognitive learning. Since the participants cannot readily utilize

the previously learned motor sequences on novel SG conditions, the transfer-related performance gains occur due to a trajectory-independent learning component. The two experiments are not independent because the same participants employed the same KM. Therefore, the improvements suggest that the KM-specific internal model is acquired and transferred from the Single-SG condition to facilitate efficient trajectory planning in the Mixed-SG condition. The acquisition of an internal model involves learning the KM relations between the possible cursor movements and the keypress buttons. This internal model is employed while planning the trajectories to generate a new sequence of keypresses that can be executed to solve a novel SG condition. Therefore, the cognitive component in GST is a form of the trajectory-independent learning process and it involves the acquisition of a KM-specific internal model.

The participants were able to employ the learned KM (from the Single-SG condition) to plan trajectories to the goal position with minimal failed attempts, as evident from a significant decrease in the error rate from 2.857 in the Single-SG condition to 0.690 in the Mixed-SG condition. The error rates denote the average number of error trials attempted to complete each successful trial. The fraction of participants who performed the first trial without any errors increased from 21% in the Single-SG condition to 69% in the Mixed-SG condition. Moreover, a qualitative examination of the evolution of trajectories in the early and late phases of the Mixed-SG condition suggests the role of a KM-specific learning component in GST. It is apparent from Figure 2.5 that the participants employed the learned KM-specific internal model to improvise on non-optimal trajectories in the early phase. Thus, the late phase is characterized by optimal trajectory planning and increased trajectory density.

This account of transfer of learning is also corroborated by improvements in other behavioral measures such as the mean execution time and average reward score. The mean execution time decreased from 3,743 to 2,807 ms as the participants were able to quickly plan trajectories using the acquired KM on novel SG conditions. The mean reward score also improved from 20.952 in the Single-SG condition to 67.976 in the Mixed-SG condition. In summary, these results suggest that the participants are faster, more accurate, and quickly discover the optimal trajectory in the Mixed-SG condition as compared to the Single-SG condition. It suggests a key contribution of transfer of the acquired key-map from the Single-SG condition to the Mixed-SG condition.

In addition, we observed a significant effect of practice on the reaction time in Single-SG and Mixed-SG conditions (see Figures 2.2 B, 2.4 B). This result is rather intriguing because no improvements in reaction time were expected in line with the previous findings (Fermin et al., 2010). The reaction time denotes the latency of the first keypress, reflecting the time cost of pre-planning the whole trajectory from the start to the goal position. A steady decrease in reaction time implies that the participants become more adept at using the previously acquired KM-specific internal model to plan trajectories with practice. The reaction time trend provides additional evidence for the involvement of cognitive learning in GST.

The Single-SG condition involved a rotation trial. One possible way to complete the rotation trial efficiently, would be to execute the learned sequence of keypresses after performing a mental rotation of the KM and SG positions to "undo" the rotation. If participants employed this strategy, we would

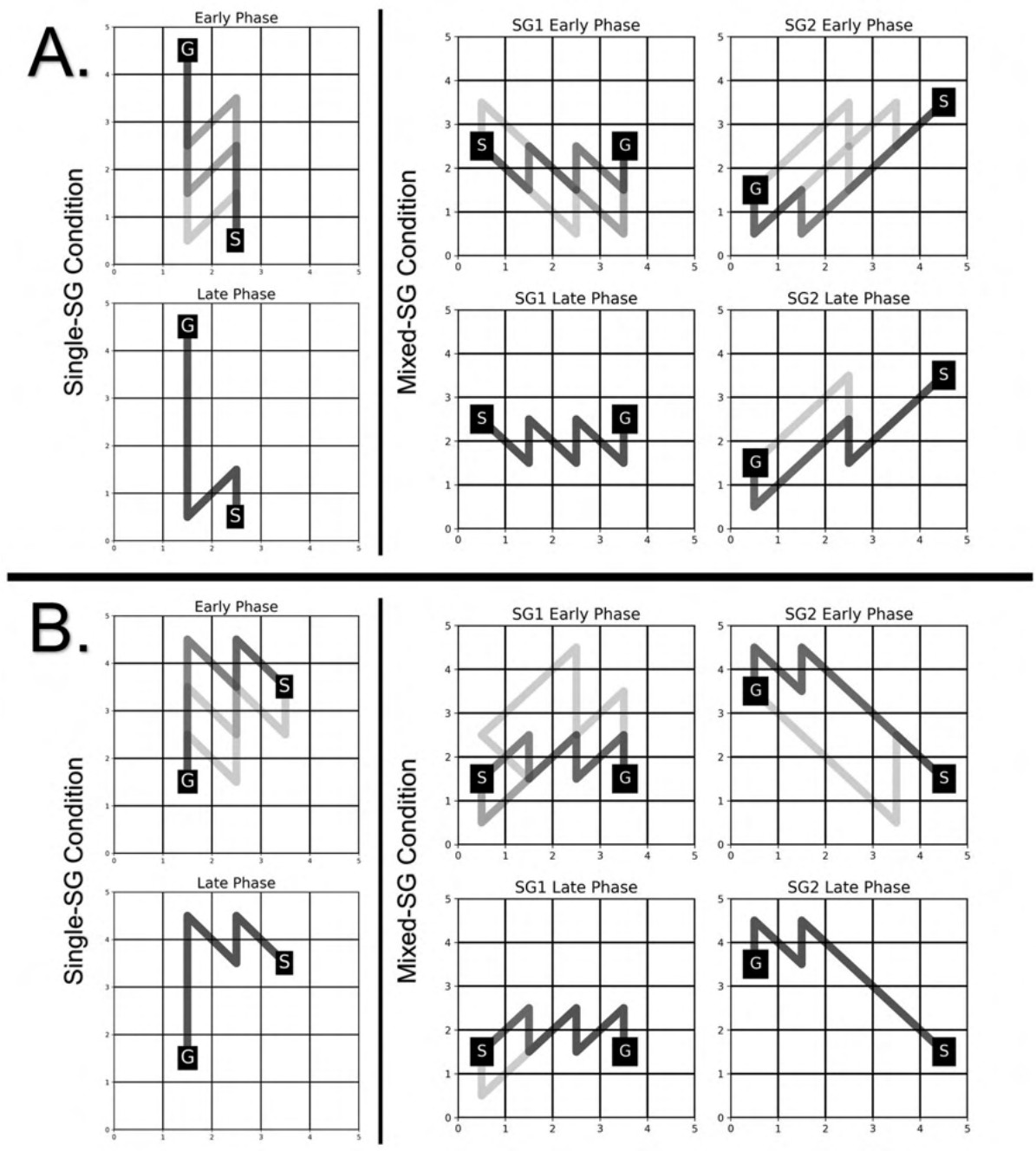

Figure 2.5 Evolution of trajectories in Single-SG and Mixed-SG condition in two representative participants—MD (A) and AS (B). Participants MD and AS are assigned key-maps KM1 and KM2, respectively. The comparison of trajectories in early vs. late phase is shown. The early and late phase correspond to the first and last five successful trials, respectively, in each condition. A darker trajectory shade denotes more frequented trajectory.

anticipate that the reaction times increase but not the execution times. However, we observed an increase in execution time and reaction time even on multiple attempts on the rotation trial (see Figure 2.3). This indicates that the participants may have attempted the rotation trial as a novel KM-SG condition. Consequently, the execution time increased due to the additional time cost of planning trajectories using a novel KM. The performance degradation is also evident from other behavioral measures such as reward score and error rates. We further examined the differences in the trajectories traversed in the normal and rotation condition (see Figure 2.6). We observed many qualitative differences between trajectories traversed in normal and rotation conditions, irrespective of the number of attempts on the rotation trial. Overall, the results in the rotation trials suggest that the trajectory-specific motor program learned in the normal condition could not be transferred to the rotation condition successfully.

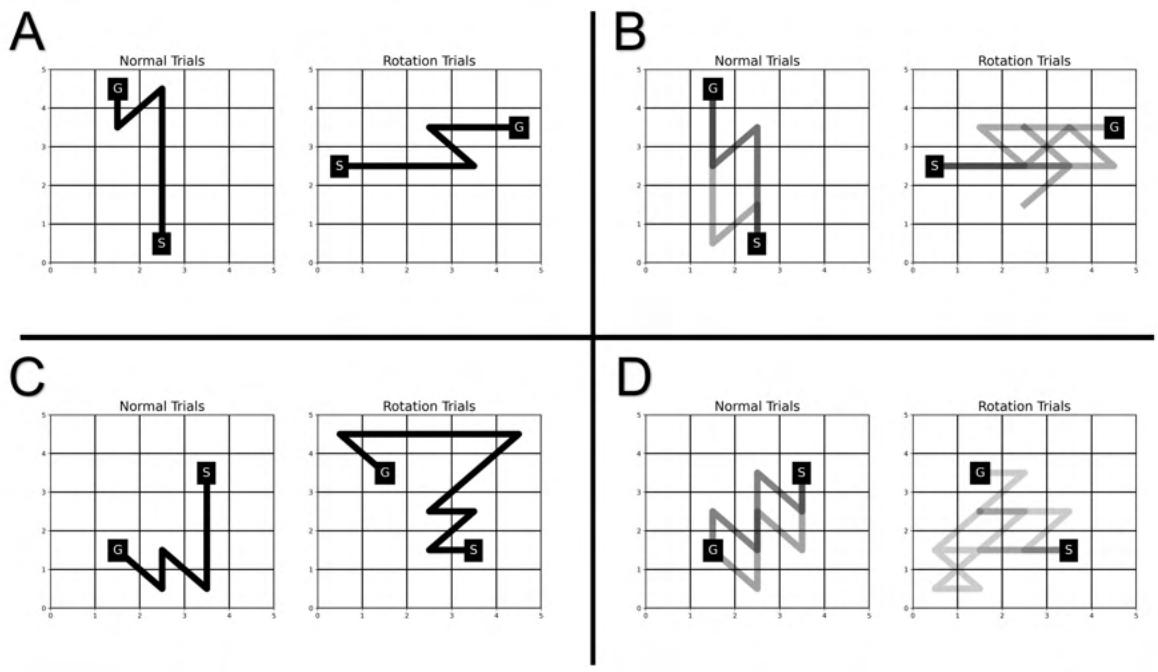

Figure 2.6 Comparison of trajectories traversed during normal and rotation trials in Experiment-1 for four representative participants—JK (A), LM (B), RS (C), and CP (D). The trajectories for all the rotation trials are plotted for each participant. The number of trajectories plotted for the normal condition is matched with that from the rotation condition. The number of trajectories (or trials) plotted for participants JK, LM, RS, and CP is 1, 3, 1, and 5, respectively. A darker trajectory shade denotes more frequented trajectory.

2.4.3 Theoretical Perspectives on Internally-Guided Sequencing

Fermin et al. (2010) provide evidence for model-based action planning in GST by demonstrating that the participants benefit from previously learned state transition models (or KM) if an additional delay is given before the start of the movements. For learned KM, such a delay would favor model-based plan-

ning using internal simulations of sequential action selection. In our case, such acquisition of the internal model is evidenced by the ability to efficiently navigate in a randomly-ordered mixed-SG condition where trajectory/sequence-specific learning due to visuomotor associations is minimized. While Fermin and colleagues established the progressive nature of learning in different stages based on model-based and model-free action selection strategies, we examined the behavior in GST in terms of cognitive-motor dichotomy in internally-guided sequencing. We provide evidence to establish the role of cognitive as well as motor learning in GST. Numerous studies have tried to reconcile the computational (model-based vs. model-free) and behavioral (cognitive vs. motor) perspectives in understanding the distinct, parallel processes involved in motor learning (Dezfouli and Balleine, 2012; Keele et al., 2003; McDougle and Taylor, 2019; Savalia et al., 2016; Wolpert et al., 2011; Wolpert and Landy, 2012). In line with Fermin et al. (2010), we show evidence for the role of the cognitive learning process as part of goal-directed, model-based action planning in GST. The implications of our findings are two-fold—while establishing the role of cognitive and motor processes in the non-trivial planning and sequence execution in GST, we call for a renewed interest in understanding a class of practical, internally-guided motor sequence learning tasks. In sum, sequencing behavior in GST involves both general motor learning and acquisition of an internal model. Learning the association between the movements and the corresponding keypresses allows for the acquisition of sequence as participants learn to "react" appropriately and efficiently to the motor intention or plan as circumscribed by the KM-specific internal model. General motor learning results in quick and efficient performance with repeated execution of finger movements in response to visual cues. The performance improvements due to motor learning may be driven by motor chunking. In GST-like internally-guided tasks, practice-driven motor learning is constrained by a goal-directed internal plan of sequential actions. The internal model in GST involves the acquisition of a general structure or organization which guides the sequential order of keypress execution. A salient feature of such paradigms is that motor learning is influenced by the structure and organization of practiced sequences. It is guided internally and not externally imposed. Consequently, internally-guided paradigms involve developing internal representations for both the response-effect mappings for KM and the sequence of keypresses (trajectory) to reach the goal. These internal representations are subject-specific even when the participants were using the same KM on a similar SG condition. Our account is again in line with the previous studies on the ideomotor framework of voluntary action control. The action-effect (or R-E) bindings emerge during action planning, integrating components of the forward and inverse models of motor control (Nattkemper et al., 2010; Ziessler et al., 2004).

GST involves the role of interleaved cognitive and motor learning components. This parallel trajectory planning and motor learning induce a natural duality in the task. We speculate the role of working memory and visuospatial attention in GST. The task involves divided attention where information such as KM and the current trajectory is actively maintained online in the working memory. In contrast, SG information in the visual buffer helps in directing the cursor towards the goal position. The executive control inhibits the natural tendency of executing a response to generate the appropriate sequence of keypresses, given the constraints of the KM-specific internal plan. The early practice phase would be

characterized by high attentional and cognitive demands as the participants learn response-effect mapping for the KM. With trial and error, as they learn to move the cursor towards the goal, the visuospatial attention and working memory are actively engaged to strategize navigation to the goal position. Further practice allows for optimizing the trajectories to reach the goal position in an optimal number of moves. Once an optimal path is discovered in the late practice phase, the performance improvements occur predominantly due to motor learning. We speculate that motor chunking characterizes automatic and habitual control with reduced attention.

Grid-sailing task is a simple canonical paradigm that does not require any complicated experimental setting, yet it offers rich insights into the planning and sequencing behavior. Unlike other discrete sequencing tasks such as SRT, DSP, or $m \times n$ task, GST involves learning self-generated motor sequences. In GST, the sequential keypresses are not guided by an external series of stimuli but are instead self-initiated by a KM-specific internal model. The behavior in GST can be organized into the "planning" and "executing" phases. These distinct phases enable natural dissociation of cognitive and motor strategies involved in internally-guided sequence learning. This is a unique and helpful characteristic of GST that can be leveraged to investigate the role of different learning processes involved in internally-guided sequencing. The cognitive phase in GST can be distinctly associated with acquisition of the trajectory-independent and KM-specific internal model, which is employed while navigating the grid. The learned KM could also enable a selective transfer of the learned model to other tasks where the KM is compatible. Therefore, GST can also be used to study skill transfer and related behavioral phenomena. Moreover, the GST task paradigm affords variations in different aspects. The GST instances can differ in various factors such as grid-size, start-goal (SG) positions used, KMs associated with the task, and the number of cursor movement directions. Owing to many possible variations in the GST paradigm, the instances cover a broad spectrum of grid-navigation tasks that vary across aspects such as the difficulty of solving, execution time required, and cognitive effort demanded—providing reasonable experimental control that is necessary to study different factors involved in sequence learning tasks.

In this chapter, we aimed to identify the learning processes involved in GST. Therefore, the results largely involved analysis of cumulative behavioral measures such as execution time and reward. In the next chapter, we aim to investigate the nature of motor learning to understand the underlying mechanisms of dexterous sequential finger movements. To this end, we identify temporal patterns in individual keypress times and understand how these patterns evolve to facilitate efficient sequence execution.

Chapter 3

Chunking in Motor Skill Learning [1]

3.1 Introduction

Motor learning is a function of two specific categories of skill acquisition–motor adaptation and motor chunking. Motor adaptation is the context-specific, error-driven acquisition of locomotor patterns, which involves adjusting to changes in sensory input or motor output characteristics (Izawa et al., 2008; Shadmehr and Mussa-Ivaldi, 1994). A practical example of motor adaptation is learning to use a mouse pointer with different sensitivities. With changes in mouse sensitivities, we adaptively learn to map our hand movements to mouse pointer movement changes. Motor chunking involves the consolidation of certain movement elements into clusters that allow efficient execution of the multi-element sequence (Sakai et al., 2003; Verwey and Eikelboom, 2003). With repeated rehearsals, the individual elements are consolidated into a sequence of quick motor actions. For example, motor chunking is observed when we learn the swift execution of keypresses while playing a videogame. Complex action sequences in the game are broken down into multiple simple components (e.g., "Jump-forward-and-Shoot" or "Move-right-Duck-Jump-forward").

Much of the early interest in motor chunking focused on how the repeated execution of visuomotor sequences leads to overall performance improvement. Studies have shown that the inter-response intervals within certain sub-sequences decrease with practice compared to those in-between these sub-sequences (Bo and Seidler, 2009; Kennerley et al., 2004; Verwey et al., 2009). It leads to the emergence of distinct clusters of motor movements (called motor chunks), which facilitates efficient sequence execution (Lashley, 1951; Rosenbaum et al., 1983; Zeigler and Gallistel, 1981). The chunks segment the motor sequences into smaller representational clusters, reflecting integrated sequence representations (Kennerley et al., 2004; Verwey et al., 2009). With substantial practice, the temporal patterns become more prominent and the chunks can be identified more distinctively (Sakai et al., 2003; Verwey and Eikelboom, 2003; Wymbs et al., 2012). The benefit of such segmentation of the action sequences is that it reduces memory load during sequence execution (Bo and Seidler, 2009; Logan, 2018; Ramkumar

[1] This chapter is a slightly modified version of our publication **Motor Chunking in Internally Guided Sequencing**; Bera, K., Shukla, A., & Bapi, R. S. (2021) in *Brain Sciences*.

et al., 2016; Thalmann et al., 2019; Yamaguchi and Logan, 2014). Each motor chunk is stored as a single memory representation and the entire chunk is loaded at once into the motor buffer for execution. This results in "automatic" control of sequential movements with reduced cognitive demand (Abrahamse et al., 2013). Previous studies have also shown that motor chunking is not merely an effect of rhythm consolidation in sequential motor tasks. The time or memory resource constraints can also inhibit motor chunk loading, leading to performance degradation (Verwey, 1996; Verwey and Dronkert, 1996). In the explicit domain, motor chunking has been shown in paradigms such as discrete sequence production task (DSP) and $m \times n$ task (Abrahamse et al., 2013; Bapi et al., 2000; Povel and Collard, 1982; Restle and Burnside, 1972; Robertson, 2007; Sakai et al., 2003; Verwey and Eikelboom, 2003). Moreover, some studies employing the serial reaction task (SRT) have also provided evidence for cluster pattern of motor sequence performance during implicit learning of visuomotor sequences (Curran and Keele, 1993; Koch and Hoffmann, 2000; Nissen and Bullemer, 1987; Stadler, 1993).

While the emergence of similar cluster patterns has been observed in a variety of motor sequence learning tasks, not many studies have examined the role of motor chunking in internally-guided sequencing. Our present study investigates the role of chunking in practice-driven performance improvements in internally-guided sequencing. We hypothesize the role of chunking in motor learning in GST. The task required participants to discover an optimal path to the goal position (via a sub-goal) using the learned KM. They executed the same trajectory in all the subsequent trials to consolidate the learning. First, we show overall learning, as reflected in improved execution times in successful trials with practice. Then, we examine the underlying temporal patterns of keypress response times to show that the sequential keypresses are organized in subsequences or chunks, facilitating efficient behavior. We show how these clusters consolidate into fewer, larger chunks with practice in internally-guided sequencing. The significance of our study lies in probing the role of chunking during motor learning in real-life paradigms. The externally-guided paradigms, such as SRT or DSP, involve "passive" motor learning behaviors, which are guided by external stimuli and so findings from these studies cannot be generalized to most practical motor tasks. Using canonical paradigms, such as grid-navigation, our study highlights how motor learning can be investigated in more practical tasks.

3.2 Methods

The repeated sequential execution of keypresses in grid-navigation tasks amounts to sequence learning in an internally-guided fashion. These tasks involve navigating a cursor from the start position to the goal position on the grid. The possible cursor movements are associated with particular keyboard buttons in a one-to-one correspondence. Each individual path or trajectory from start to goal position constitutes a novel sequence of keypresses. The optimal trajectory to the goal is dependent on other task specifications such as possible agent movements and reward schema. To complete the trial successfully, the participants can choose to reach the goal position using any possible optimal trajectories. The repeated execution of these trajectories results in learning a self-generated, voluntary sequence of

keypresses. This behavior during the sequential execution of motor actions provides us with rich insights into how we become increasingly proficient in internally-guided tasks. Therefore, we employ an adapted version of grid-sailing task (Fermin et al., 2010) to investigate chunking behavior in internally-guided sequencing. Moreover, GST is a simple task and so the dissociation of the sequence-specific motor learning is relatively easier when controlling for the trajectories executed. It is also a flexible canonical paradigm that can be altered on factors, such as key mapping, start-goal position, size of the grid and reward schema, to facilitate this dissociation and flexibly generate multiple variations of the task.

3.2.1 Participants

Fifteen right-handed participants (7 women and 8 men) between ages 18 to 27 years (mean = 21.53; SD = 2.79) performed the experiment for partial course credits. All participants were healthy with normal or corrected-to-normal vision. The experiment was approved by the Institute Review Board, IIIT-Hyderabad, India. The participants gave informed consent before the study. Two participants did not complete the experiment as they were unable to recall the first optimal trajectory that they traversed. The data from their attempts were excluded for all purposes. The data from the remaining thirteen participants were used for all analysis purposes.

3.2.2 Apparatus

The participants were seated on a chair facing a high-resolution 24-in computer screen placed approximately two feet away. The responses were recorded using a conventional computer keyboard. The participants used the right index, middle and ring fingers to press the numpad buttons “4”, “5” and “6”, respectively. Other keys were removed to prevent meddling in response selection. Custom-made programs were written using Python3 and PyGame (Python Game Development) for stimulus presentation and data recording.

3.2.3 Procedure

The subjects were given verbal instructions about the task rules before the session started. Each trial began with the presentation of a 10×10 grid with a red fixation on the center of the screen. On pressing the space button, after a random delay of 500–1000 ms, the trial would begin with the start position marked as a green tile, the sub-goal position marked as a red tile and the goal position marked as a blue tile. The cursor was shown as a black triangle, initially placed in the starting position. The participants were given 9 s to solve each trial, and this duration was not explicitly conveyed to them. We computed this to be an ideal adjusted trial duration based on the average length of optimal trajectories and the mean execution time in the original GST study (Fermin et al., 2010). During the trial duration, participants executed sequential keypresses to navigate the cursor from the starting position to the goal position via

the sub-goal. The sub-goal was introduced to help participants in navigating to the goal position. To complete the trial, the participants must reach the goal position only after visiting the sub-goal position. The possible cursor-movement directions were defined by the key mapping (KM) (see Figure 3.1 B). Apart from the possible number of movement directions, no other information about the key mapping was conveyed to the participants. The task required participants to explore the possible KM movement directions and the corresponding key associations by trial-and-error.

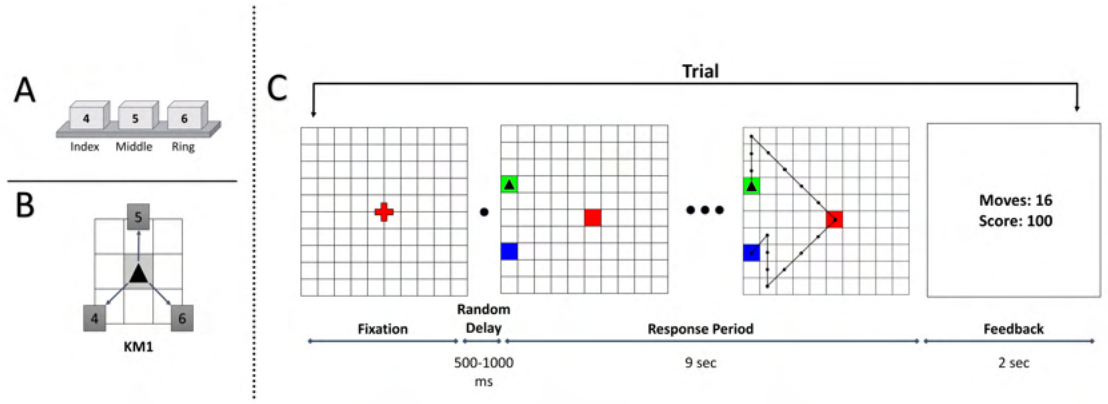

Figure 3.1 (A) Numpad keys and the respective hand fingers. (B) Key mapping (KM) used in the experiment. The marked arrows show possible movement directions. The boxed numbers indicate the numeric keys associated with the movements. (C) Task diagram: sequence of trial events. The green, red and blue tiles show the start, sub-goal and goal position, respectively. An example optimal trajectory is shown on the grid while using the KM from Fig. B

The participants were instructed to achieve a maximum score (of 100 points) while executing each trial as quickly as possible. A maximum of 100 points was awarded when the participants traversed an optimal path to reach the goal position. A minimum-steps trajectory from start to goal via the sub-goal position is considered an optimal path. Suppose a non-optimal path was traversed, a penalty of -5 points incurred for every excess move. In case the participant tried to perform an infeasible move, such as moving out of the grid, the cursor stayed there, but the action increased the move count. If the participant failed to reach the goal position in the given time duration, 0 points were awarded for that trial. At the end of each trial, the performance feedback was presented for 2 s, following which the fixation screen signaled the beginning of a new trial. In the center of the feedback screen, the performance feedback was presented as two numbers: the number of moves in the traversed trajectory and the trial reward score. A trial illustration is shown in Figure 3.1 C.

The trajectory was controlled for all subsequent trials to investigate the practice-driven performance improvements with the repeated execution of sequences. The pair of start-goal (SG) positions and the sub-goal position remained constant throughout the experiment. The participants were asked to remember the first optimal path (minimum-steps trajectory; reward score = 100) traversed on the grid. Once they discover an optimal path, the trajectory traversed in that particular trial was locked in the program. In all subsequent trials, the participant must repeatedly traverse the same path to reach the

goal and complete the trial successfully (reward score = 100). Any deviation from the locked path will award the participant zero points and the trial will be deemed unsuccessful. The experiment ran until subjects successfully performed 60 trials while repeatedly traversing the optimal path that they first discovered. The participants were given a rest block after every 20 trials to minimize the effects of muscle fatigue on task performance.

3.2.4 Behavioral Measures

The number of moves in the traversed trajectory, reward obtained, reaction time and execution time were the performance variables recorded for each trial. Individual keypress response times (RTs) were also recorded for key-level analysis purposes. Reaction time is defined as the time interval between the onset of stimuli and the first keypress. Execution time is computed as the difference between the keypress time of the last and the first response. For analysis purposes, the trials were classified into two categories (1) successful trials—trials with perfect reward score (equal to 100), and (2) error trials—trials with an imperfect reward score (not equal to 100). Only successful trials were included for all analysis purposes.

3.3 Results

3.3.1 Learning in GST

Participants performed the grid-navigation task and the behavioral measures on each trial were recorded. On average, each participant attempted 87 trials to complete 60 successful trials. The error trials constituted 30.6% of all trials. To examine how the learning evolves with GST, we plotted mean error rates across all participants. The error rate is computed as the number of error trials attempted to complete each successful trial. In Figure 3.2 A, we observe that most of the error trials occur during the initial trials of the task when the participants are learning to use the key map to find an optimal path to the goal position. The error rates drastically decrease after participants discover the first optimal trajectory.

On plotting mean execution times for successful trials across participants, we see a decreasing trend in the plot (see Figure 3.2 B). The decrease in execution times over trials can be attributed to performance improvements due to learning. With practice, the participants learned the KM and utilized it to plan an optimal trajectory to the goal position. Further performance improvements followed as they learned to execute the trajectory sequences swiftly. The law of practice effect on learning is evident by examining the initial and last trials. In the beginning, the participants took a longer mean execution time of 6247 ms (SD = 1303) to complete the first trial successfully. After substantial practice, the mean execution time on the last trial was significantly lower at 3822 ms (SD = 766). The execution time improvements suggest that the participants learned to navigate and execute sequential keypresses efficiently with practice.

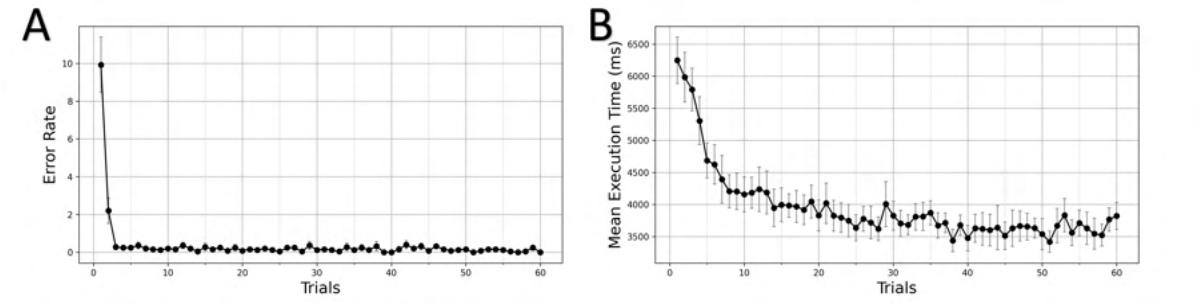

Figure 3.2 (A) Learning in grid-sailing task (GST): the error rates decrease as participants (N = 13) discover the optimal trajectory. (B) Trial-by-trial course of performance improvement in execution time across participants (N = 13) in successful trials. The bars on plot data-points denote standard error.

A non-parametric Friedman test of differences among repeated measures (within-subjects) rendered a significant effect of trials on the execution time $(\chi^2(59) = 291.198, p < 0.001)$.

3.3.2 Motor Chunking in GST

The decrease in execution time over trials is illustrative of the performance improvement due to skill learning. The improvement in execution time was observed even when participants repeatedly executed the locked optimal trajectory (i.e., reward = 100). While the reward certainly guided the planning of motor actions during initial trials by providing feedback on the optimality of the trajectory, it becomes inconsequential to motor performance once the optimal trajectory is discovered. At this point, the reward only indicates whether the same optimal trajectory was followed in the subsequent trials. Such binarized reward feedback incentivizes the participants to repeatedly execute the same internally-guided trajectory throughout the remaining experiment. This enables us to probe how spontaneous grouping structure emerges in sequence execution as the motor performance is fine-tuned and optimized. We hypothesize the role of motor chunking in practice-driven performance improvements. We identify the motor chunks based on changes in temporal patterns of keypress RTs.

3.3.2.1 Identifying Chunk Patterns from Keypress RTs

The chunks were identified using Wilcoxon signed-rank tests between successive keypress response times (RTs). Studies have shown that the initial element in a chunk typically exhibits a slower behavior because of the RT cost of initializing and loading the motor chunk (Barnhoorn et al., 2019; Izawa et al., 2008; Newell, 1991). Therefore, the keypress RT for the first element of the chunk is significantly different from the next element. In agreement with this argument, if the n^{th} and $(n+1)^{st}$ keypress RTs are significantly different and the $(n+1)^{st}$ keypress RT is less than n^{th} keypress RT, both will belong to the same chunk. Additionally, in case the $(n+1)^{st}$ keypress RT is significantly higher than the n^{th} keypress RT, both keypresses will not belong to the same chunk. All successive elements with non-significant keypress RT differences belong to the same chunk. To be considered as a chunk, the segment

should have at least two keypress elements. Therefore, the three operating rules to identify chunks are (i) the initial element in a chunk is typically characterized by a significantly higher RT than the following keypresses; (ii) only successive elements with a statistically insignificant difference or monotonically decreasing keypress RTs are appended to the current chunk; (iii) a significant increase in keypress RT denotes the beginning of a new chunk. Figure 3.3 shows motor chunks (marked by brackets) in four representative subjects in early and late practice phases. The first ten trials (trials 1–10) belong to the early practice phase, whereas the last ten trials (trials 50–60) belong to the late practice phase.

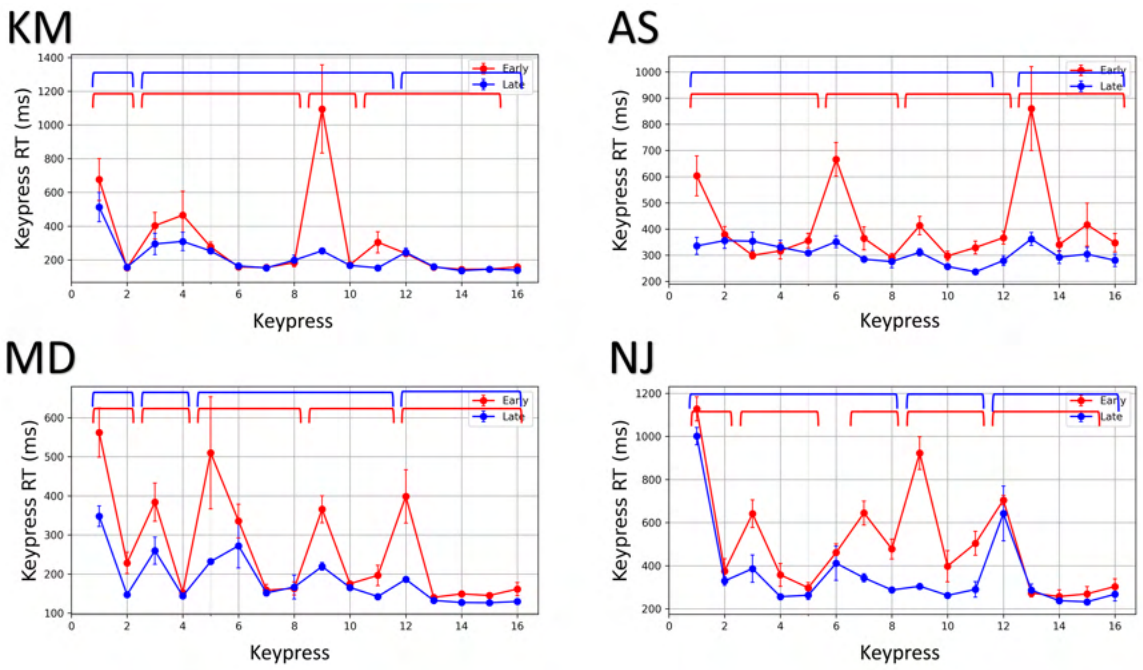

Figure 3.3 Re-organization of chunks with practice. Average keypress response time (RT) comparison plots for early (trials 1–10) and late (trials 50–60) phase in four representative subjects (K.M., A.S., M.D. and N.J.). The early and late phase RTs are plotted in red and blue, respectively. The brackets in red and blue on the top of each plot denote chunks in the early and late phases.

3.3.2.2 Re-Organization of Action Sequences with Practice

To corroborate our findings, we also examine the emergence and re-organization of motor chunks from the early practice phase to the late practice phase. Previous studies have shown that the motor chunking in sequence learning tasks evolves with practice-induced changes in temporal patterns of execution (Newell, 1991; Willingham, 1998). To check if the performance in the early phase is different from that in the late phase, we computed the mean keypress RTs in each phase by averaging all the keypress RTs (1–16) in each phase (10 trials) across all the subjects (N = 13). A paired t-test between mean keypress RTs for early (mean = 358 ms; SD = 68) and late (mean = 257 ms; SD = 49) phase showed a significant difference $(t(12) = 11.435, p < 0.001)$.

As the temporal patterns of sequence execution dynamically evolve with practice, the chunks re-organize with repeated concatenation and segmentation. Table 3.1 records changes in chunk features such as the number of chunks formed and the average length of chunks over the course of practice in four representative subjects. For example, subject N.J. executed the entire sequence in five segments during the early phase (red) and three segments during the late phase (blue). The average chunk length increased from 2.80 to 5.33. The area between the two lines (red and blue) in Figure 3.3 indicates performance gains with the re-organization of the chunks. A general increasing and decreasing trend were found across all participants from the early to late phase for the average length of the chunk and the number of chunks, respectively (see Figure 3.4). The average number of chunks across subjects in the early and late phase were 3.923 and 3.154, respectively. As normality assumptions were violated, we used a Wilcoxon signed-rank test, which suggests that the decrease in the number of chunks was significant ($df = 12, Z = 2.5, p = 0.026$). The average chunk length across subjects increased from 4.153 in the early phase to 5.263 in the late phase. A Wilcoxon signed-rank test suggests that the increase in chunk length was also significant ($df = 12, Z = 6.5, p = 0.018$).

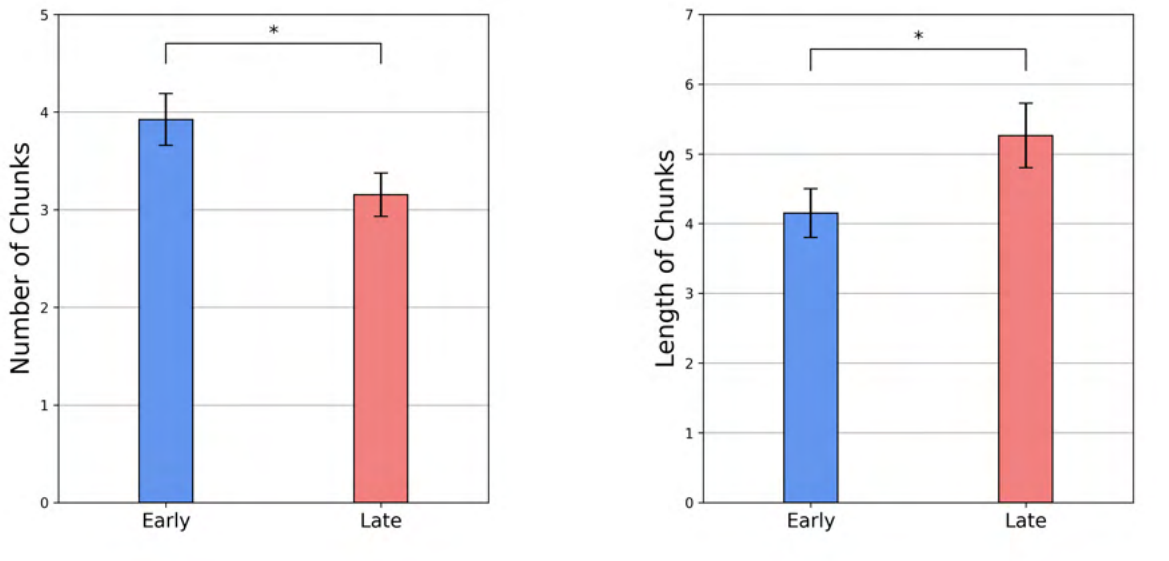

Figure 3.4 Evolution of chunking behavior with practice. The number of chunks significantly decreased, whereas the length of the chunks significantly increased from the early phase to the late phase. The bars denote standard error. * $p < 0.05$.

3.4 Discussion

Motor chunking has been extensively studied in externally-guided tasks in both implicit and explicit domains. Not many studies have investigated chunking in internally-guided sequencing tasks. To the best of our knowledge, this is the first study investigating motor chunking in internally-guided

Subject	Early Phase		Late Phase	
	No. of Chunks	Avg. Length of Chunks	No. of Chunks	Avg. Length of Chunks
K.M.	4	3.75	3	5.33
A.S.	4	4.00	2	7.50
M.D.	5	3.20	4	4.00
N.J.	5	2.80	3	5.33

Table 3.1 Re-organization of chunks with practice. A comparison of the number of chunks and length of chunks in the early and late phase for four representative subjects, K.M., A.S., M.D. and N.J.

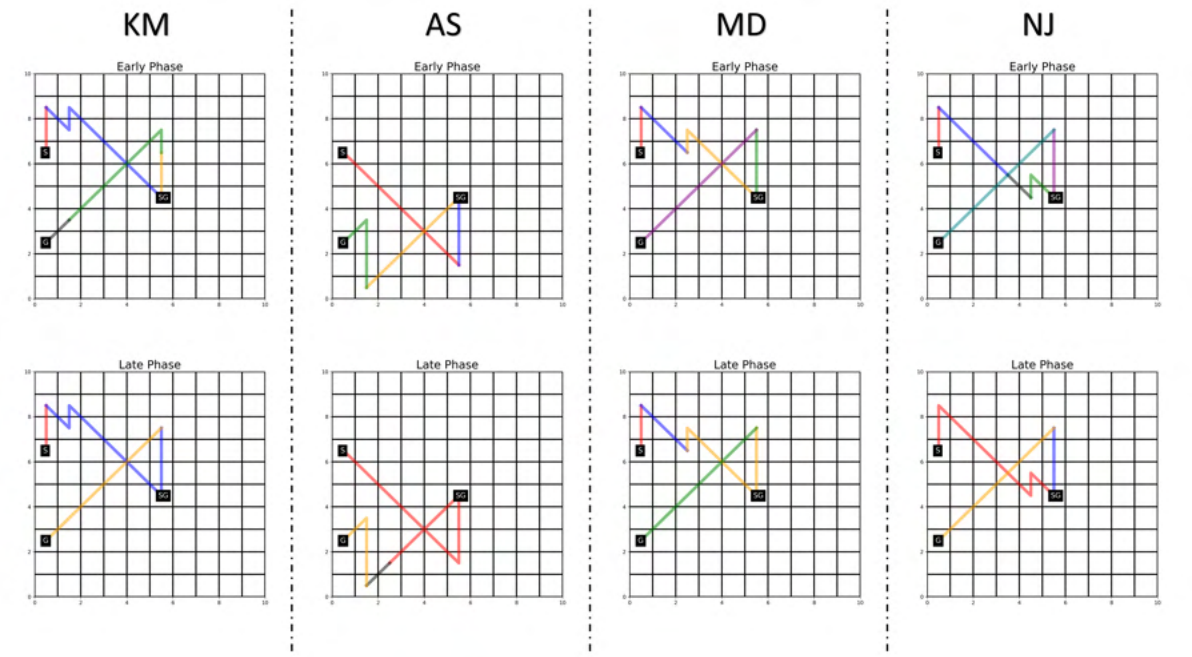

Figure 3.5 Re-organization of chunks with practice. Chunks are overlayed the trajectories for early (trials 1-10) and late (trials 50-60) phase in four representative subjects (K.M., A.S., M.D., N.J.). Each color denotes individual chunks. The moves marked in black do not belong to any chunk.

paradigms. Using grid-navigation as an exemplar paradigm, we hypothesized the role of motor chunking in practice-driven performance improvements in internally-guided sequencing. First, we analyzed the effect of trials on execution time to show trajectory-specific motor learning in GST. Then, we analyzed the temporal patterns of keypress RTs to provide evidence for motor chunking. Distinct clusters of swiftly executed successive elements emerge with practice. We found that the keypress RTs are different in the early and late practice phase. We observed a significant improvement in keypress RTs in the late phase due to chunk consolidation. We further showed how the chunk characteristics, such as chunk length and number, evolve to facilitate efficient execution. The number of chunks decreased as the length of chunks increased from the early practice phase to the late practice phase (see Table 3.1 and Figure 3.4). With practice, smaller chunks coalesce to form bigger chunks to promote simpler integrated sequence representations that can be executed efficiently. We also observed that the chunk boundaries did not simply correspond to switching between the keys, reiterating that chunking is not merely functional (see Figure 3.5). We also found that in the early phase, the participants were slower on the sub-goal, and hence, the sub-goal marked a chunk boundary. As the chunks evolved with practice, the sub-goal was consequently executed as a medial keypress within the chunk for most participants. It is indicative of the re-organization of the chunks, which happens to facilitate efficient execution. The chunking behavior was found universally across all subjects. The chunking pattern in the sequence was individual or subject-specific.

Our study corroborates findings from other studies on chunking and internally-guided tasks. The learning in GST is a function of both cognitive and motor components. While cognitive learning relates to the ability to navigate the grid using the acquired key map, motor learning involves acquiring skillful sequencing behavior. The motor learning in GST is characterized by the improving dexterity of executing the sequences repeatedly. In line with previous work in externally-guided tasks, we show that the sequence learning in GST-like internally-guided tasks is facilitated by segmenting the sequence into motor chunks (Schmidt et al., 2019; Shadmehr and Mussa-Ivaldi, 1994). Our findings also corroborate with the computational account of learning proposed in (Verwey and Dronkert, 1996). The chunking-driven efficiency can be attributed to the model-free, memory-based strategy that the participants use to reproduce pre-learned action sequences after extensive practice. Previous studies have identified chunking as a cost-effective learning strategy that reduces overall computational complexity while maintaining efficiency (Haibach et al., 2018). Neuro-imaging studies investigating GST and cued-sequence production tasks provide evidence for the shared neural underpinnings of motor-memory guided actions and chunking (segmentation and concatenation) processes. The supplementary motor area, putamen and anterior cerebellum areas are involved in GST in the late phase when the actions are habitual, automatized and driven by model-free learning (Fermin et al., 2016). The studies have shown the role of sensorimotor putamen, frontoparietal network and pre-supplementary motor area during chunking behavior in sequencing tasks (Barnhoorn et al., 2019; Clegg et al., 1998). Given these complementary findings, future work can investigate if a computational framework of chunking can be conceptualized with the joint modeling of response times and decisions in internally-guided sequencing tasks.

While our study replicates previous findings from the chunking literature on a less constrained, internally-guided sequencing paradigm, it is important to situate the behavioral phenomena in chunking in the context of the learning processes involved in internally-guided paradigms. In contrast with externally-guided actions, internally-guided or intention-based actions involve higher-order cognitive processes such as planning, memory and decision-making. Therefore, it is crucial to understand whether internally-guided and externally-guided sequence learning have similar underlying mechanisms in light of this differentiation. The sequence learning in externally-guided paradigms has long been explained with multiple single-level accounts (see Abrahamse et al. (2010) for a review) of response location associations, perceptual and response effect learning. Subsequent studies can probe if similar learning mechanisms are at play during internally-guided sequencing. The internally-guided tasks will allow us to disentangle and understand the contributions of perceptual and response–effect learning. In line with some previous results (Ziessler and Nattkemper, 2001), we speculate that serial learning in internally-guided tasks involves voluntary action control mechanisms characterized by the acquisition of response–effect contingencies. In the ideomotor framework of action and perception, voluntary action goals have been shown to influence internally-guided sequencing (Hoffmann et al., 2001; Hommel, 2003; Hommel et al., 2001). Based on previous neuroimaging studies (Jueptner et al., 1996; Jueptner, 1998; van Donkelaar et al., 1999), we anticipate the role of basal ganglia, pre-supplementary motor cortex and dorsolateral prefrontal cortex in chunking in internally-guided sequencing tasks.

In the previous chapters, we have identified the processes underlying learning in GST and further established the role of motor chunking. However, these behavioral experiments did not inform us about how the learning progresses in GST. In the next chapter, we examine the stage-wise progression of learning using an inter-manual transfer task. We discuss how early and late practice phases are characterized by differential engagement of the two learning processes or representations.

Chapter 4

Inter-manual Transfer of Motor Skills

4.1 Introduction

In our everyday experiences, skill learning and skill transfer seem to be related phenomena. The most common of them is the case of inter-manual transfer of skills. For example, it is difficult to perform tasks such as writing, drawing and chopping using the non-dominant hand compared to the dominant hand. It is easy and efficient to perform certain tasks on the 'trained' dominant hand. This can be explained based on a dichotomic framework of cognitive-motor and effector-dependent-effector-independent learning.

Studies on the Grid-Sailing Task (Bera et al., 2021a) and other paradigms (Doya, 2000; Ghilardi et al., 2009; Penhune and Steele, 2012) have shown the role of cognitive and motor components in skill learning. The cognitive component is concerned with the acquisition of spatial-temporal order of movements in the sequence. The motor component involves optimization of the fine motor movements, which facilitate efficient sequence execution. In addition to the cognitive-motor dichotomy, studies have shown evidence that the sequence learning employs two independent representations and processing (Hikosaka et al., 1999; Keele et al., 1995; Kumar et al., 2020; Verwey and Clegg, 2005; Verwey and Wright, 2004). Neuro-imaging studies have suggested two different neural circuits underlying these two independent processes that code the sequence representations (Bapi et al., 2006; Hikosaka et al., 2002; Perez et al., 2007). The learning in visual-spatial coordinates is effector-independent, whereas the learning in motor coordinates is effector-specific (Bapi et al., 2000). The effector-specific motor learning is usually processed implicitly and, therefore, slowly acquired. It requires minimum attentional and working memory demands (Hikosaka et al., 1999). On the other hand, the cognitive component is often linked to the acquisition of abstract sequence mechanisms (van Mier and Petersen, 2006). It is fast developing, accurate in space but slow. It is typically characterized by significant cognitive demands (Hikosaka et al., 1999). While there has been evidence (Shea et al., 2011) to show that effector-independent representations in visual-spatial coordinates encode movement sequences more effectively even after substantial practice, Bapi et al. (2000) showed that the time courses of acquisition during learning are different for both the processes. The reliance on visual-spatial representation gradually decreases with practice. Other studies have also confirmed the findings that the sequence becomes increasingly effector-dependent with

practice (Kovacs et al., 2009; Park and Shea, 2003; Verwey and Wright, 2004). The practice induces rapid motor skill improvements that are non-transferable irrespective of the number of acquisition trials (Boutin et al., 2012). Moreover, studies have also shown that transfer is symmetric when visual-spatial coordinates are reinstated after the acquisition of sequence (Kovacs et al., 2009; Panzer et al., 2009). However, the transfer was not symmetric when the motor coordinates were reinstated.

While studies (Thomas et al., 2012) have investigated such accounts of learning in externally-guided sequencing, not much is known about the progressive nature of learning in internally-guided sequencing. In this experiment, we employ a transfer task to probe the stage-wise transitions of the underlying learning processes during the acquisition of motor sequences. We aim to understand what kind of sequence knowledge is acquired during different phases of learning. We hypothesize that learning in GST involves a transition from a dominant role of the cognitive component (effector-independent representation) in the early phase to the motor component (effector-dependent representation) in the late phase. The participants performed GST on the same SG position in three sessions on consecutive days. Each session started with a normal block where the participants used the dominant hand to perform the task. The normal block was followed by a transfer block where the participants used the non-dominant hand. Across days, the transfer block was introduced after varying amounts of practice on the normal block. To empirically test the cognitive to motor 'switch' hypothesis, we compare the performance measures on normal and transfer blocks across days. We anticipate increasing differences in performance between the normal and transfer blocks as a function of practice.

4.2 Methods

4.2.1 Participants

Fourteen participants (11 men and 3 women) between ages 19 to 25 years (mean = 22.36; SD = 1.84) performed the experiment for partial course credits. All the participants were healthy, right-handed individuals with normal or corrected-to-normal vision. The experiment was approved by the Institute Review Board, IIIT-Hyderabad, India. The participants gave informed consent before the study. Not all participants completed the experiments as a few of them dropped out of the sessions on Day-2 and Day-3.

4.2.2 Apparatus

The participants were seated on a chair facing a high-resolution 24-in computer screen placed approximately two feet away. The responses were recorded using a conventional computer keyboard. The participants used the numpad buttons "4", "5" and "6" to execute the movements. Other keys were removed to prevent meddling in response selection. Custom-made programs were written using Python3 and PyGame (Python Game Development) for stimulus presentation and data recording.

4.2.3 Procedure

The subjects were given verbal instructions about the task rules before the session started. Each trial began with the presentation of a 10×10 grid with a red fixation on the center of the screen. On pressing the space button, after a random delay of 500–1000 ms, the trial would begin with the start position marked as a green tile, the sub-goal position marked as a red tile and the goal position marked as a blue tile. The cursor was shown as a black triangle, initially placed in the starting position. The participants were given 9 s to solve each trial, and this duration was not explicitly conveyed to them. During the trial duration, participants executed sequential keypresses to navigate the cursor from the starting position to the goal position via the sub-goal. The sub-goal was introduced to help participants in navigating to the goal position. To complete the trial, the participants must reach the goal position only after visiting the sub-goal position. The possible cursor-movement directions were defined by the key mapping (KM) (see Figure 4.1 A). Apart from the possible number of movement directions, no other information about the key mapping was conveyed to the participants. The task required participants to explore the possible KM movement directions and the corresponding key associations by trial-and-error. The participants were randomly assigned one of the two possible SG conditions used in the experiment (see Figure 4.1 B). Thus, 5 participants performed GST on SG1 and 9 participants performed GST on SG2.

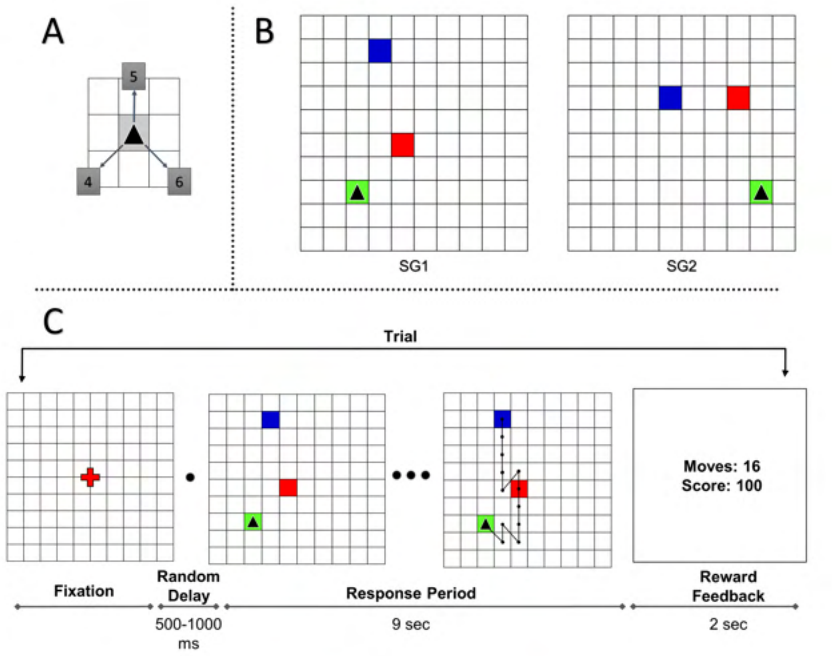

Figure 4.1 (A) Key mapping (KM) used in the experiment. The marked arrows show possible movement directions. The boxed numbers indicate the numeric keys associated with the movements. (B) The start-goal (SG) conditions used in the experiment. The green, red and blue tiles show the start, sub-goal and goal position, respectively. (C) Task diagram: sequence of trial events. An example optimal trajectory is shown on the grid while using the KM from Fig. A

The participants were instructed to achieve a maximum score (of 100 points) while executing each trial as quickly as possible. A maximum of 100 points was awarded when the participants traversed an optimal path to reach the goal position. A minimum-steps trajectory from start to goal via the sub-goal position is considered an optimal path. If the participant traversed a non-optimal path or failed to reach the goal position, 0 points were awarded for that trial. In case the participant tried to perform an infeasible move, such as moving out of the grid, the cursor stayed there, but the action increased the move count. At the end of each trial, the performance feedback was presented for 2 s, following which the fixation screen signaled the beginning of a new trial. In the center of the feedback screen, the performance feedback was presented as two numbers: the number of moves in the traversed trajectory and the trial reward score. A trial illustration is shown in Figure 4.1 C.

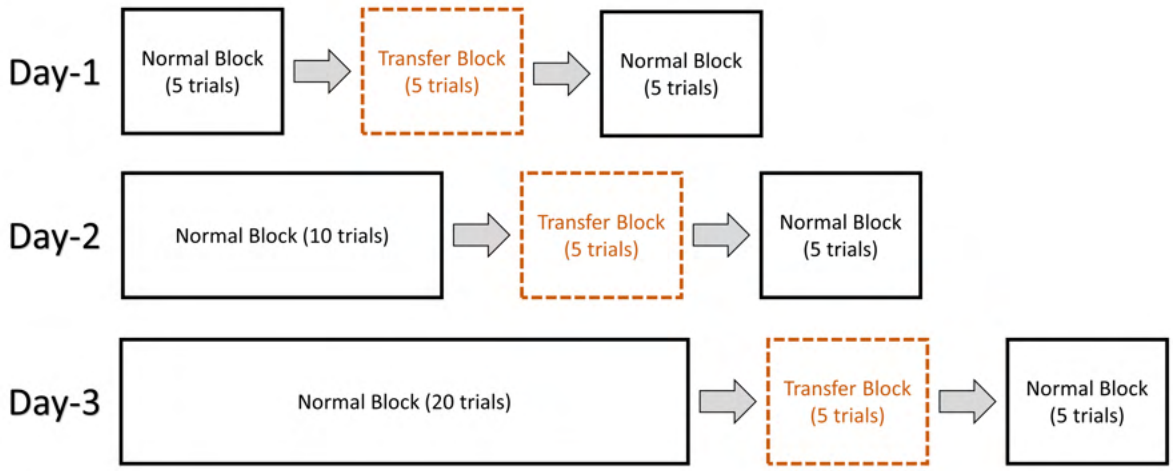

Figure 4.2 Transfer task experiment design. The normal blocks (black) are performed with the dominant hand whereas the transfer block (brown) are performed with the non-dominant hand. On all the three days, the participants first performed GST on normal block followed by the transfer block. The transfer block is introduced after varying amount of practice on the normal block.

The transfer experiment involved three task sessions conducted on consecutive days. The task involved dominant to non-dominant transfer to probe the progressive nature of learning in GST. The session involved performing GST on the normal block and then on a transfer block on all three days. The normal block was performed using the dominant hand, whereas the transfer block was performed using the non-dominant hand. The task design had interleaved normal and transfer blocks. The length of the first normal block on the three days varied (5 successful trials on Day-1, 10 successful trials on Day-2 and 20 successful trials on Day-3). The first normal block was followed by the transfer block

(5 successful trials). The transfer block was followed by the second normal block (5 successful trials). Figure 4.2 shows the transfer experiment design.

4.2.4 Behavioral Measures

The number of moves in the traversed trajectory, reward obtained, reaction time and execution time were the performance variables recorded for each trial. Individual keypress response times (RTs) were also recorded for key-level analysis purposes. Reaction time is defined as the time interval between the onset of stimuli and the first keypress. Execution time is computed as the difference between the keypress time of the last and the first response. For analysis purposes, the trials were classified into two categories (1) successful trials—trials with perfect reward score (equal to 100), and (2) error trials—trials with an imperfect reward score (not equal to 100). Only successful trials were included for all analysis purposes.

4.3 Results

The execution time on the successful trials was plotted for all three days (see Figure 4.3). As the participants learn to execute the sequence efficiently, we see a decreasing trend in the execution time on all three days. On Day-1, the execution time improved from 5166 ms (SD = 1299) in the first successful trial to 3558 ms (SD = 1318) in the last successful trial of the first normal block. A non-parametric Friedman test of differences among repeated measures (within-subjects) rendered a significant effect of trials on the execution time on Day-1 ($\chi^2(4) = 18.80, p < 0.001$). Similarly, on Day-2, the execution time improved from 3039 ms (SD = 571) in the first successful trial to 2703 ms (SD = 824) in the last successful trial of the first normal block. A Friedman test on execution time indicated a significant effect of trials on Day-2 ($\chi^2(9) = 20.821, p = 0.013$). On Day-3, the execution time improved from 2817 ms (SD = 493) in the first successful trial to 2399 ms (SD = 327) in the 20$^{\text{th}}$ successful trial. However, a Friedman test on execution suggested a non-significant effect of trials on Day-3 ($\chi^2(19) = 26.704, p = 0.112$).

To probe performance improvements over the span of days, we compared the execution time on the initial trials of all three days. The execution time decreased from 5166 ms (SD = 1299) in the initial trial of Day-1 to 3039 ms (SD = 571) in the initial trial of Day-2. A Wilcoxon signed-rank test revealed that this decrease in execution time was significant ($df = 12, Z = 91, p < 0.001$). The execution time further decreased to 2817 ms (SD = 493) in the initial trial of Day-3. A Wilcoxon signed-rank test suggested that this decrease was non-significant ($df = 6, Z = 24, p < 0.109$).

We also observed practice-related performance improvements in the transfer block across days. The execution time for the first transfer trial improved from 3538 ms (SD = 737) on Day-1 to 3115 ms (SD = 926) on Day-2 and further to 2778 ms (SD = 670) on Day-3. A Wilcoxon signed-rank test suggested that the difference in execution time on Day-1 and Day-2 was non-significant ($df = 12, Z = 73, p < 0.057$).

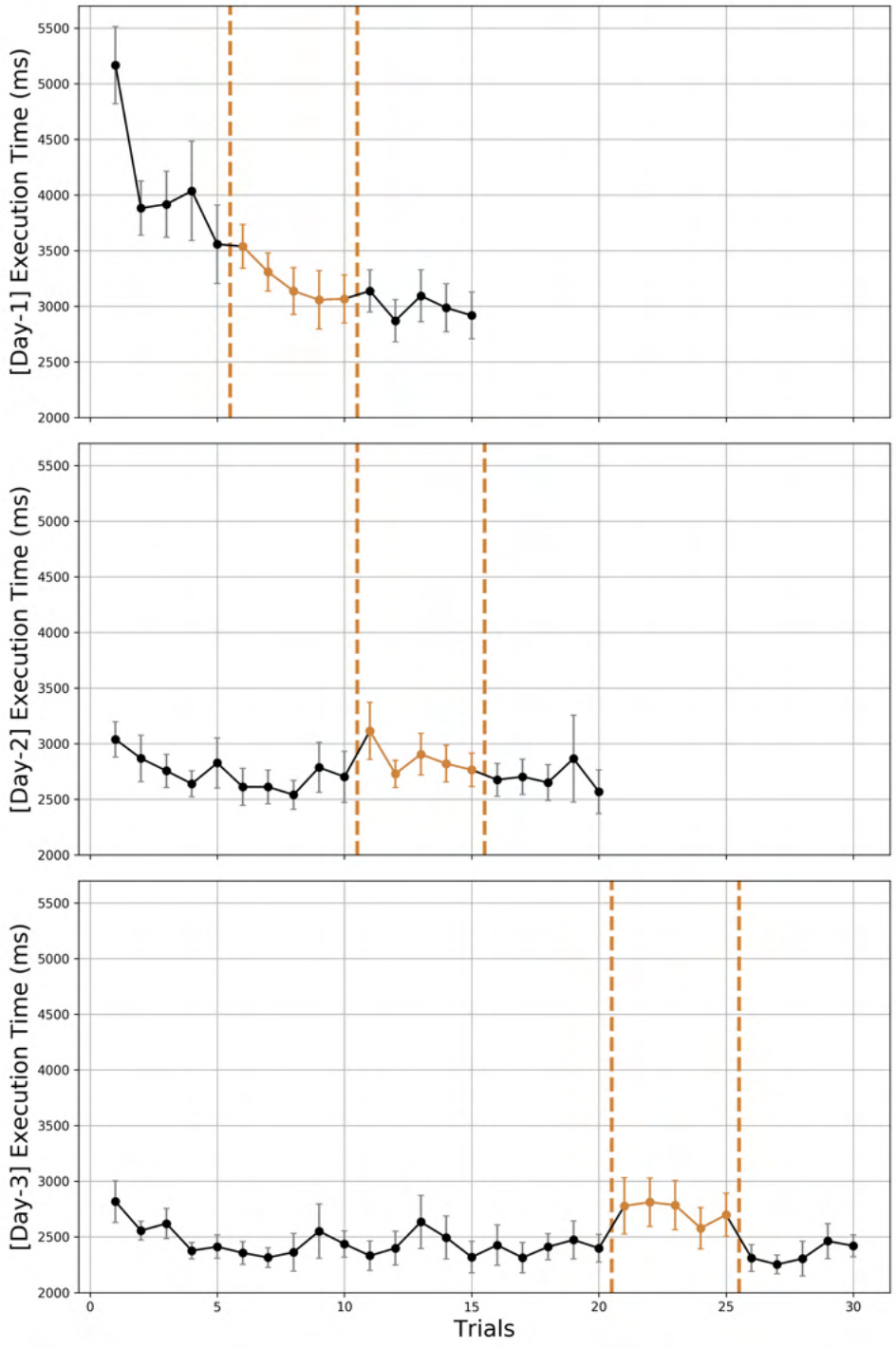

Figure 4.3 Mean execution time in the transfer task across three days. The brown dotted lines denote transfer block start/end and the brown datapoints denote execution time on transfer block.

Similarly, the difference in execution time on Day-2 and Day-3 was also non-significant ($df = 6, Z = 16, p = 0.813$).

4.3.1 Performance comparison between the normal and transfer block

We compared the mean execution time on the transfer block and the second normal block. On Day-1, the execution time increased from 3001 ms (SD = 658) on the normal block to 3221 ms (SD = 683) on the transfer block. A Wilcoxon signed-rank test indicated that the difference was non-significant ($df = 13, Z = 83, p = 0.058$). On Day-2, the execution time in the normal block was 2693 ms (SD = 706). The execution time increased to 2867 ms (SD = 550) on the transfer block. A Wilcoxon signed-rank test indicated that the difference was non-significant ($df = 12, Z = 65, p = 0.191$). Similarly, on Day-3, the execution time increased from 2350 ms (SD = 301) on the normal block to 2730 ms (SD = 533) on the transfer block. A Wilcoxon signed-rank test suggested that the difference was significant ($df = 6, Z = 28, p = 0.016$).

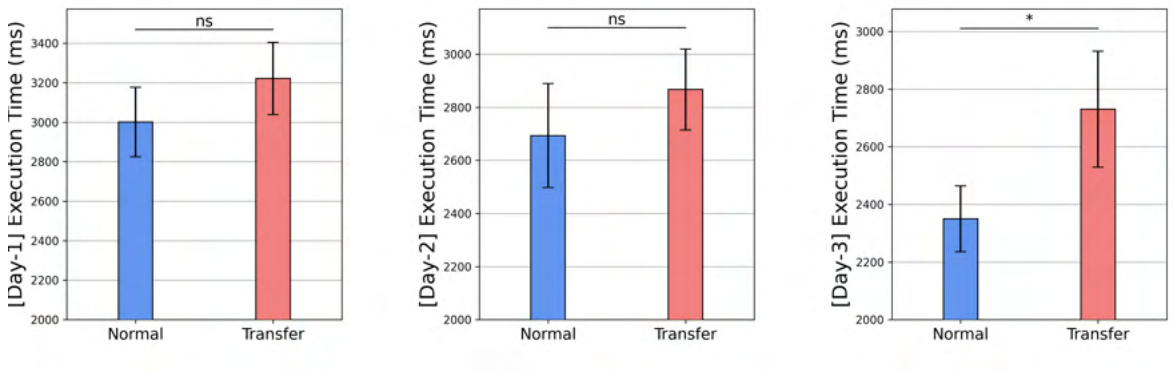

Figure 4.4 Comparison of performance on the normal and transfer blocks. Mean execution time is plotted for the transfer block and the (second) normal block for each day. ns non-significant; $* p < 0.05$

4.4 Discussion

We performed an inter-manual transfer experiment to gain insights into the nature of representations acquired with learning progression. The experiment design involved introducing a transfer block after varying amounts of practice on the normal block across the three days. Thus, the experiment involved a switch from the dominant hand to the non-dominant hand. The rationale for employing dominant to non-dominant switch was to minimize practice-related transfer effects. Studies have shown that sequence learning is asymmetrically transferred during inter-manual transfer tasks - the transfer is better from the non-dominant to the dominant hand (Hicks, 1974; Parlow and Kinsbourne, 1990; Sakai et al., 2003; Taylor and Heilman, 1980).

In our transfer experiment, the SG condition remained the same across all three days on both conditions. The optimal trajectory to the goal position (the sequence of keypresses) remained the same in both conditions. However, in the transfer condition, the sequence of motor responses (finger movements) differed from the normal condition because participants were using different hands in both conditions. Since the untrained, non-dominant hand is used on the transfer block, the sequence execution is facilitated solely by the learning in the abstract, visual-spatial domain. Therefore, the performance in the transfer block is only a function of the effector-independent learning. While we observed practice-related performance improvements in transfer blocks across days, the decrease in execution time was non-significant.

A trend in performance differences between the normal and transfer block would suggest the extent to which the sequence learning is restricted to effector-independent learning. The effector-independent learning occurs at a more abstract level, such as response selection or response locations (Keele et al., 1995; Willingham et al., 2000). If learning relies predominantly on the effector-independent knowledge, then we do not anticipate significant differences in performances while using either hand. On the other hand, if the learning is primarily effector-dependent, then the performance on the trained hand would be better than that on the untrained hand.

Our findings suggest that the participants become progressively better at performing the task in normal conditions. With substantial practice, an optimized motor program is learned that facilitates quick and efficient execution of the sequence. Given that the acquisition of such a motor program is effector-specific, the performance on the transfer block does not benefit from this motor learning. Therefore, we observed increasing differences in performance on the normal and transfer block across the days. During the early practice phase, the learning is primarily effector-independent, and therefore, we did not observe significant differences on Day-1 and Day-2. The late practice phase involves the dominant role of effector-dependent representation, and therefore, we observed that the trained, dominant hand is efficient at performing the task than the untrained, non-dominant hand. In short, sequence learning relies on effector-independent representations in the early practice phase and on effector-dependent representations in the late practice phase.

In line with previous studies, we show a transition from effector-independent learning to effector-dependent learning during sequencing behavior (Bapi et al., 2000; Hikosaka et al., 2002). Our findings validate a phase-wise switch from the cognitive to the motor learning processes. Moreover, our findings corroborate with Fermin et al. (2010)'s account of learning. They hypothesized the role of multiple controllers in sequential action selection in GST. They show how these controllers take a dominant role in sequencing based on the progress of learning. Once the participants map the actions (keypresses) to their results (movements), action selection likely involves a dominant role of planning for successful goal-reaching. The later practice phase is characterized by quick and efficient execution of motor responses. Therefore, the learning transitions from relying on abstract, response-based representations to more effector-dependent representations with practice. This is evidenced by the increasing differences in performances on normal and transfer blocks across the three days in our experiment. The follow-

ing chapter presents a computational equivalence of cognitive and motor learning and discusses how it might relate to effector-independent and effector-dependent representations in learning.

The current and previous chapters presented an empirical investigation of the typical behavioral phenomenon in GST. In the next section of the thesis, we examine internally-guided sequencing from a computational perspective. We first situate the behavioral phenomenon in GST in a dual process account of learning. Then we investigate if a reinforcement learning based computational account of sequence learning gives additional insights into internally-guided sequencing.

Chapter 5

A Computational Account of Learning [1]

5.1 Reinforcement Learning

Reinforcement Learning (RL) is an unsupervised way of training an agent to take actions in the environment while maximizing some reward objective (see Figure 5.1). Using RL, the agent learns from the environment on its own. There is no involvement of a supervisor that guides the learning. Unlike other classes of machine learning algorithms, RL does not involve any training over data. The only input to the RL agent is the dynamic feedback it receives while interacting with the environment. Using this feedback, the agent iteratively learns to make sequential decisions that maximize its reward function. RL can be used as a general computational framework for solving numerous real-world tasks. The applications range from training an artificial agent to play chess, operating a robot in a factory, to building an autonomous driving vehicle.

To use RL in the problem at hand, we must re-structure the problem in a particular mathematical formulation called a Markov Decision Process (MDP). The MDP is based on the fundamental assumption of Markov property. The Markov assumption states that given the present, the future is independent of the past. It means that the future course of action is only dependent on the current state of the environment. The actions taken in the past do not matter. Brownian motion is an example of a random process that follows Markov property. The MDP can be mathematically formalized in a tuple as $< S; A; T; R; \gamma >$. Here, S is the set of states in the environment; A is the finite set of possible actions that the agent can take in the environment; T is the transition probabilities matrix; R is the reward function defined over the states and γ is the discount factor that denotes discounting for the temporally-distant rewards. The transition probabilities matrix, or simply the transition table, is of the form $T(s; a; s')$ which denotes the probability of transitioning from state s to state s' on taking action a. The reward function takes the form $R(s; a; s')$ which indicates the immediate reward received by the

[1] This chapter is a modified version of our publication **Value-of-Information based Arbitration between Model-based and Model-free Control (Oral)**; Bera, K., Mandilwar, Y., & Raju, B. (2019) in *Sixth Annual Conference of the Cognitive Science Society, Goa, IN*.

agent on taking action a in state s while leading to a transition to state s'. The transition probabilities and the reward function together are referred to as the 'model' of the environment.

The mathematical objective in RL is to solve for optimal policy or optimal values. The policy ($\pi(s) : \mathbb{S} \rightarrow \mathbb{A}$) is analogous to a look-up table that maps a state to an action. The optimal policy is the best policy yielding highest expected cumulative rewards for the agent. The optimal policy can be determined from the state values ($V(s)$). The state values denote the 'good-ness' or utility of being in a particular state. The state value is the expected cumulative reward for the agent in a given state. Mathematically, $\forall s \in \mathbb{S}$ it is expressed as

$$V(s) = \mathbb{E}[\sum_{i=1}^{T} \gamma^{i-1} r_i],$$

where T is the length of the episode. The state-action value or the Q-value ($Q : \mathbb{S} \times \mathbb{A} \rightarrow \mathbb{R}$) denotes the expected cumulative reward while taking a particular action in the given state. An optimal policy implicitly implies knowledge of optimal state values and optimal state action values. In other words, $\forall s \in \mathbb{S}$,

$$V^*(s) = \max_{\pi} V^{\pi}(s) = \max_{a} Q^*(s, a),$$

where $V^{\pi}(s)$ denotes value function for a given policy π, $V^*(s)$ denotes the optimal state value and $Q^*(s, a)$ denotes optimal state-action value.

The fundamental principle underlying RL is learning by trial-and-error. As the agent interacts with the world, it receives feedback. This feedback is used to learn and determine the next best action that needs to be taken. For example, if the agent receives a very large positive reward for taking a particular action in the given state, it will consider it as a beneficial one. The agent will try to execute the same action when it is in that state to maximize the rewards. On the other hand, a negative reward or penalty will refrain the agent from taking the chosen action again.

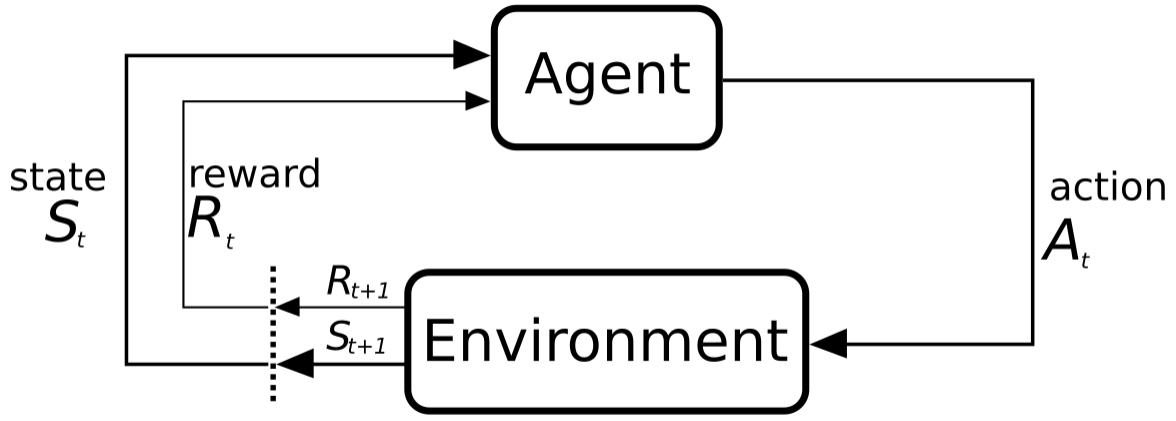

Figure 5.1 Reinforcement learning: learning by trial-and-error (Sutton and Barto, 2018).

5.1.1 Model-free RL

As the name suggests, model-free RL does not require the use of transition probabilities or the reward function. Model-free RL estimates the value functions directly based on the interactions with the environment. It involves an iterative update of the state-action values by trial and error from the agent's experiences. It is a suitable RL solution when the model of the environment is unknown or the environment is too complex. Model-free RL is commonly implemented using temporal difference (TD) learning. In TD learning, the agent's value estimates are updated at each decision step once the reward is obtained on taking some action. It can learn from incomplete episodes or sequences. The new value estimate is obtained by updating the older estimates with a scaled target error. The target error is computed as the difference between target utility of the state and the older estimate. The scaling factor is called the learning rate (denoted by α). Here we consider two types of TD learning methods that estimate the optimal state-action values (see Figure 5.2). Model-free RL is computationally inexpensive but sample inefficient as compared to the Model-based RL.

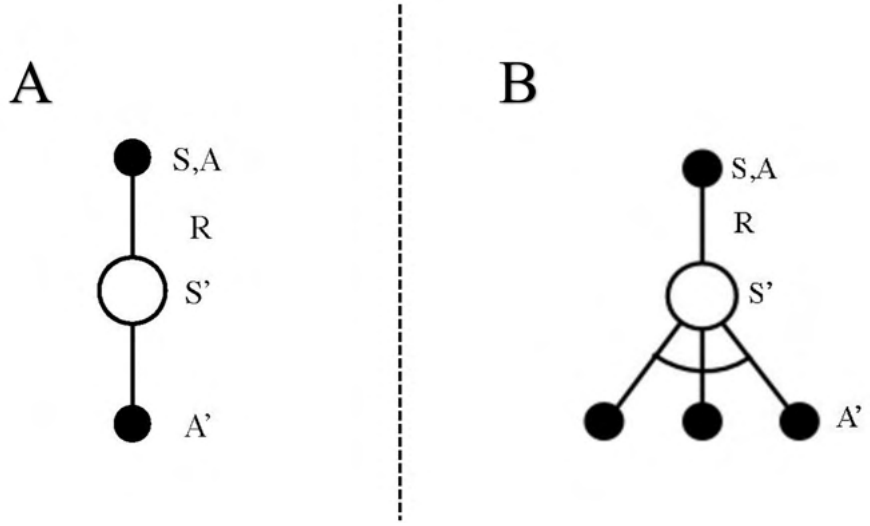

Figure 5.2 (A) SARSA and (B) Q-Learning backup diagram (Sutton and Barto, 2018).

5.1.1.1 SARSA

SARSA stands for State-Action-Reward-State-Action. It is an on-policy temporal-difference learning algorithm, that is, it learns the same policy which is used to decide what actions to take. The target policy and the behavior policy is the same. The update rule for SARSA is

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

5.1.1.2 Q-Learning

Q-Learning is an off-policy temporal-difference learning algorithm (Watkins and Dayan, 1992) almost similar to SARSA. The only difference is that the learned policy is different from the policy which is used to guide actions. The update rule for Q-Learning is

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha[r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)]$$

5.1.2 Model-based RL

Model-based RL is used when the internal model of the environment is known. It aims to construct a model of the environment based on the actual interactions. Using the acquired model, it involves planning to learn the optimal policy. Thus, Model-based RL involves learning the optimal state values with reduced interactions with the environment. It leverages the functional and structural representation of the environment to quickly learn the optimal course of actions (see Figure 5.3). The acquired model is used to simulate plans and estimate expected cumulative rewards for a given state. Model-based RL is very sample efficient as compared to the Model-free RL but the planning is computationally expensive.

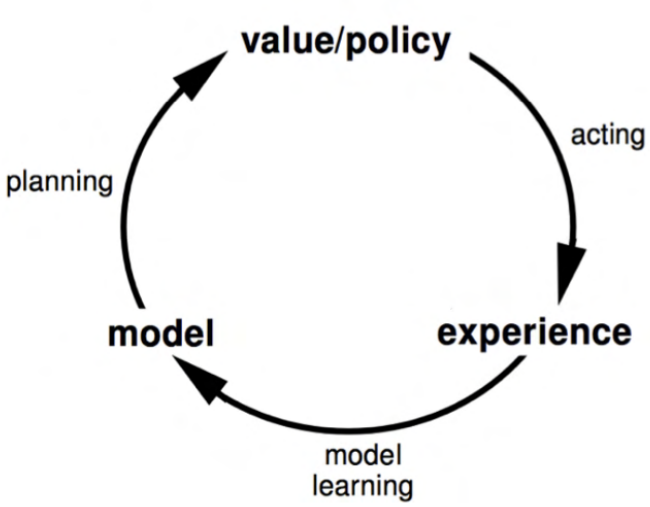

Figure 5.3 Model-based Reinforcement Learning (Sutton and Barto, 2018).

5.1.2.1 Depth-Limited Search

We implemented model-based RL using depth-limited search algorithm. The value function for a state represents discounted cumulative reward that the agent is expected to obtain in that state. It can be computed using the Bellman update equations (Sutton, 1991; Sutton and Barto, 2018). Therefore, the

optimal state values are computed as

$$V^*(s) = \max_a Q^*(s,a)$$

$$Q^*(s,a) = \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V^*(s')]$$

$$V^*(s) = \max_a \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V^*(s')]$$

The state value $V(s)$ denotes average cumulative discounted reward, weighted by the probability of transitioning to the next state s' on taking action a in state s. In order to compute the state values $V(s')$, the depth-limited search unrolls the model-based tree of state, action and resulting state to a pre-specified depth (see Figure 5.4). The value of $V(s')$ is obtained by adding the expected returns of actions up to roll-out depth and the cached state value at the deepest level of the tree. The updates are then propagated to the root node of the search tree.

At each step, as the agent interacts with the environment, the transition table is updated to reflect the probability of transitioning from state s to another state s' on taking action a. The state prediction error (SPE) is computed as

$$\delta_{SPE} = 1 - T(s,a,s')$$

The transition probability is updated as

$$T(s,a,s') = T(s,a,s') + \eta\delta_{SPE},$$

where η is a free parameter controlling the learning rate (Gläscher et al., 2010). The state transition probabilities for all the other states s'' other than s' are updated as

$$T(s,a,s'') = T(s,a,s'')(1-\eta)$$

While depth-limited search is not as computationally efficient as a value iteration or policy, the studies have shown that such a model is a more biologically plausible mechanism of forward planning (Keramati et al., 2016; Mushiake et al., 2006).

5.2 Dual Process Account of Skill Learning

The sequence learning has been explained in a dual-process framework of two parallel learning processes. The typical behavioral phenomenon in skill learning has long been explained based on a functional dichotomy of the involved processes. For example, while some studies (Bapi et al., 2000; Hikosaka et al., 1999) situate the parallel learning in visual-spatial and motor coordinates, other studies (Fermin et al., 2010; Huang et al., 2011; Lee and Schweighofer, 2009; Smith et al., 2006; Wolpert et al., 2011; Wolpert and Landy, 2012) provide an account of learning based on the complementary roles of slow and fast learning processes. Other studies have investigated the role of procedural versus

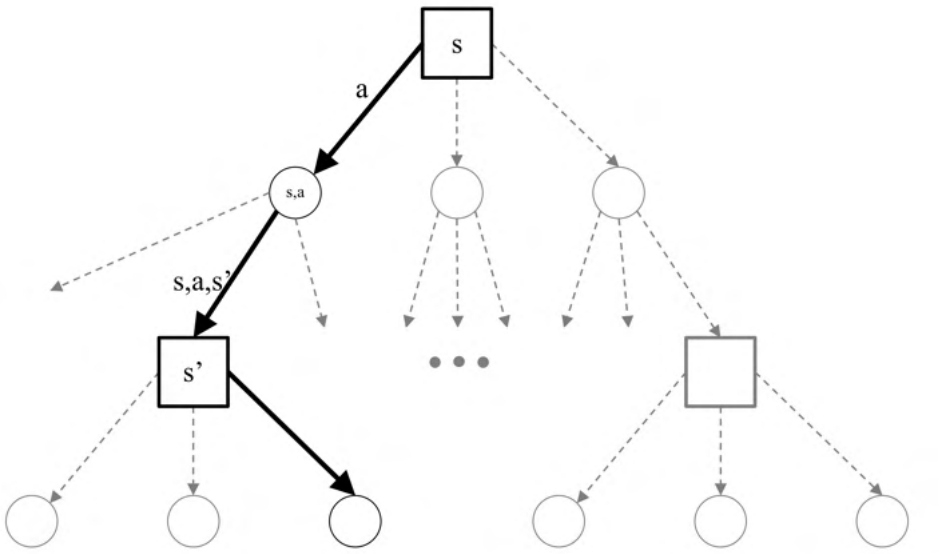

Figure 5.4 Model-based tree as internal representation of the environment

declarative memory and implicit versus explicit learning (Keele et al., 2003). Theoretical frameworks have also been proposed to provide a unifying account of these dichotomies (Savalia et al., 2016).

Skill learning involves learning a sequence of actions, the performance on which typically improves when executed multiple times. In this context, the skill learning problem, when formulated as an MDP, can be solved using reinforcement learning. Studies have show that motor skill learning involves sequential decision making (Chen et al., 2017; Huang et al., 2011; Wolpert and Landy, 2012). Haith and Krakauer (2013) argued that motor learning involves both model-based and model-free mechanisms of control. They describe that the brain implements a forward model, which is updated using sensory prediction errors. The forward model consists of an internal representation used to predict future states of the motor system (Shadmehr and Krakauer, 2008). The indirect updates to the forward model enable the control policy to adapt to perturbations and plan movements. On the other hand, model-free learning leads to the acquisition of optimal control policy with repeated error-driven adjustments. Other studies have also suggested complementary roles of dual learning processes and associated neuronal pathways in motor control (Doya, 1999, 2000; Galea et al., 2011; Shadmehr and Krakauer, 2008).

5.2.1 Habitual Learning as Model-free RL

Habitual learning results from experience-dependent repetitive behaviors (Graybiel, 2008). Habitual learning is exhibited through the reinforcement of the associations between the stimuli and the re-

sponses. Such instrumental learning is computationally inexpensive but insensitive to situations where the learned actions result in undesired outcomes (Adams and Dickinson, 1981). The habitual control is characterized by computational efficiency and automaticity of execution with reduced cognitive demands (Dayan, 2009). Studies have identified the neural underpinnings of habitual learning. It is shown that dorsolateral striatum is involved in instrumental learning (Yin and Knowlton, 2006; Yin et al., 2004). The habitual control is, therefore, computationally modeled using model-free RL (Daw et al., 2005).

5.2.2 Goal-directed Learning as Model-based RL

Studies have shown empirical evidence for goal-directed planning behaviors in action control (Balleine and Dickinson, 1998; Balleine and O'Doherty, 2010; Tolman, 1948). Goal-directed learning involves multi-step planning to achieve the desired goal. The planning in goal-directed learning involves the acquisition of an internal representation (called a 'model') that captures the environmental dynamics. The model is used to simulate and evaluate the consequences of actions to estimate the expected outcomes of a sequence of actions. Neuro-imaging studies have confirmed the role of prefrontal cortex and anterior striatum in goal-directed action selection (Balleine and O'Doherty, 2010; O'Doherty, 2011). Using such a model, the agent learns the optimal plan of action that can be executed to achieve the objective at hand. In other words, goal-directed learning is acquired through the knowledge of the actions and the corresponding outcomes. Therefore, goal-directed learning demonstrates adaptive and flexible behavior when a change (for example, reward devaluation) is introduced in the environment. It also involves substantial computation costs with the engagement of higher-order cognitive processes (Dayan, 2009). Given the analogous properties, goal-directed control is computationally described by model-based reinforcement learning (Khamassi and Humphries, 2012).

5.3 Arbitration between Model-free and Model-based RL

Multiple behavioral and neuro-imaging studies provide support for a dual-process account of motor skill learning. From a computational perspective, the dual processes can be modeled using the principles of reinforcement learning. As described in the previous sub-sections, the habitual control can be modeled using model-free RL, whereas the goal-directed control can be modeled using model-based RL. However, an open question to investigate is how the brain arbitrates between multiple controllers. How does the behavior arise from the interactions between the dual processors? The evidence suggests that the arbitration is dependent on factors such as task complexity (McDougle and Taylor, 2019), time constraints for planning (Fermin et al., 2010) and uncertainty (Daw et al., 2005) among others. Here we outline some of the previous attempts at reconciling the complementary roles of habitual and goal-directed processes.

The study in Daw et al. (2005) proposed an arbitration scheme that selects the output of the dominant processor, which predicts the optimal action with the least uncertainty. Lee et al. (2014) proposed

a similar arbitration scheme where the degree of control exhibited by each processor is a function of the reliability of their respective predictions. Gläscher et al. (2010) represented the action values of the hybrid learner as a weighted sum of action values of model-free (SARSA) and model-based (FORWARD) learners. In another study, Keramati et al. (2016) present a 'plan-until-habit' strategy in order to balance the speed and accuracy tradeoffs with an intermediate form of control. Their model simulates the tree of future states and actions up to a certain depth and then adds the cached habitual assessment of the remaining states at the deepest level of the tree to estimate the state values for decision-making. The study in Pezzulo et al. (2013) proposes a value-of-information based arbitration in order to flexibly combine the use of model-free (Q-learning) and model-based (sequential Monte-Carlo) methods for solving a double T-maze environment.

In this work, we propose two hybrid schemes - Value-of-Information based arbitration and weighted-hybrid arbitration. Using simulations, we compare the performance of both these models to the other agents such as Q-Learning, SARSA and random. We further perform model-fits to the experimental data and identify the best-fit model.

5.3.1 Value-of-Information(VoI) based arbitration

Based on the model put forward in Pezzulo et al. (2013), we propose a VoI based arbitration that balances the speed-accuracy trade-off. The advantage of such arbitration is that the model exhibits the comparative advantage of model-based and model-free mechanisms at different stages of the learning. The model-based RL is implemented as a forward tree-based search using a depth-limited search algorithm. For all simulation and model-fitting purposes, we have set the maximum depth as 2. When the max depth is set as zero, it acts like a normal off-policy algorithm with a one-step look-ahead. If the max depth is set to some arbitrarily high value (infinity), the model behaves like a sampled value-iteration algorithm. The model-free RL is implemented using the Q-Learning algorithm.

A cost-benefit meta-control involves computation of the VoI. The VoI represents the advantage of using goal-directed planning in terms of improving the state-action estimates. The expensive goal-directed planning is only invoked if the benefits of such a computation outweigh its costs. In case the benefits are non-substantial as compared to the costs, computationally inexpensive habitual behavior is preferred.

The VoI is computed as

$$VoI(s,a) = \frac{C_{(s,a)}}{\sigma(s) + \epsilon},$$

where $C_{(s,a)}$ is the uncertainty, defined as the variance of the state-action values and $\sigma(s)$ is the std. deviation in state-action values $Q(s,.)$. We add some quantity ϵ to ensure that the denominator is non-zero. The cost of the model-based tree search in the t^{th} trial is given by

$$VoI_{threshold} = a.e^{b \times t},$$

where a and b are free parameters denoting the offset and slope of the function, respectively.

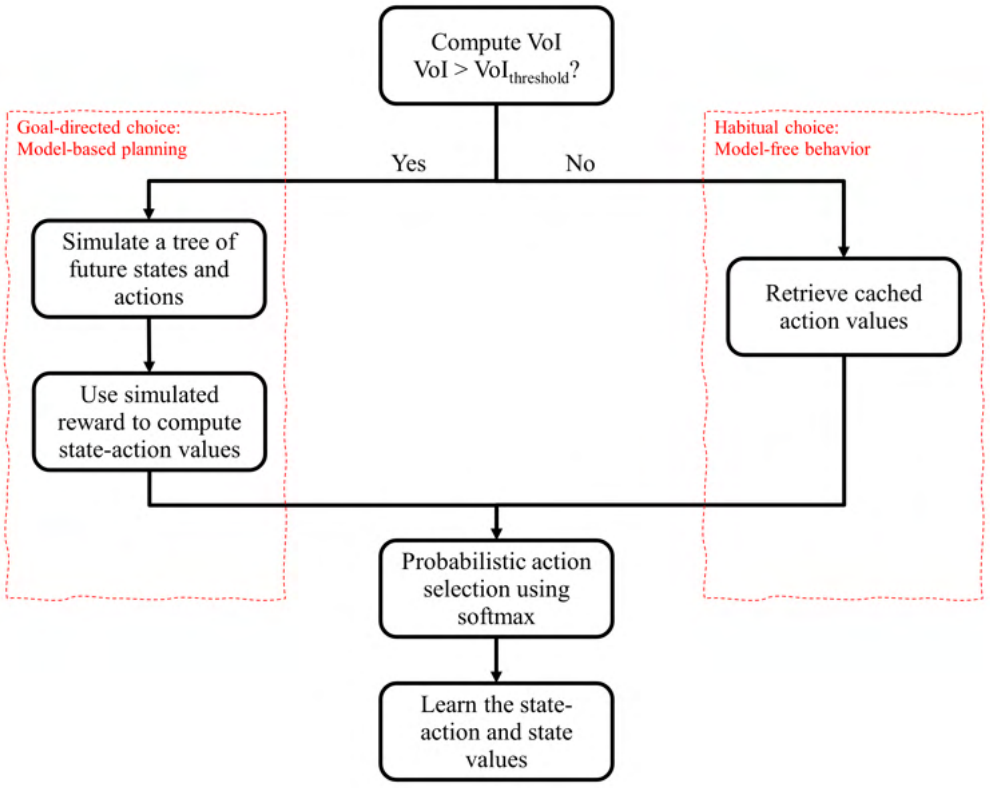

Figure 5.5 A schematic diagram of the VoI based arbitration (Adapted from Pezzulo et al. (2013)).

Therefore, the interaction with the environment occurs in either of the two stages: 'planning' and 'acting' (see Figure 5.5). If the VoI (benefit) is greater than the threshold (cost), it implies a greater uncertainty about the estimates of the state values and thus, model-based planning is employed to improve the estimates. On the other hand, if the VoI is less than the threshold, the agent is certain about the outcome of its actions in the given state and therefore, the model-free behavior is executed. The model employs a stochastic action selection using a softmax function on the state-action values. The probability of choosing action a in the state s is given by

$$P(a|s) = \frac{e^{\beta \times Q(s,a)}}{\sum_{b=1}^{n} e^{\beta \times Q(s,b)}},$$

where β is the inverse-temperature parameter, which modulates the 'greediness' in action selection. As $\beta \to 0$, the selection is random and if the $\beta \to \infty$, the selection is focused on the highest-valued action.

We tested the computational model by simulating the Mixed-SG condition from Chapter 2. Figure 5.6 shows a simulation run using the VoI based arbitration. We observed an increase in reward over the trials, suggesting that the agent was able to navigate to the goal position in an optimal number of moves. We also plotted the arbitration between the dual controllers. We counted the number of model-based evaluations and the number of model-free evaluations in each trial for all the simulation runs. The fraction of total evaluations was computed for both the controllers. We observed that model-based RL dominates the initial phase of learning. Almost 70% of the evaluations were model-based in the initial trial. This is because in the early training, the agent has not explored the environment. Thus, VoI is high and the agent invokes goal-directed planning. The model-based state-space search enables the agent to update the uncertain or unknown estimates of the state-action values. The updated estimates can then be employed to follow an optimal course of action. We observed that the model-free controller executes a smaller fraction of the total evaluations since the agent is not certain about the consequences of its actions in the environment. With practice, the fraction of model-based evaluations decreased to almost zero, whereas the fraction of model-free evaluations increased. As the agent explores the environment and learns the state-action values, the agent becomes more certain of its actions, and consequently, VoI decreases. This results in decreasing reliance on the model-based controller. The model-free controller dominates the late training phase.

5.3.2 Weighted-hybrid arbitration

Based on the proposal in Gläscher et al. (2010), we implemented a weighted-hybrid model which combines the state-action value estimates from the model-based and model-free learner. The model-free updates are implemented using the Q-Learning algorithm, whereas the model-based updates are implemented using the depth-limited search. The model-based and model-free RL maintain separate estimates of the state-action values. The values are integrated by taking a weighted sum of estimates from both the controllers. The state-action values for the weighted-hybrid learner are computed as

$$Q_{weighted-hybrid}(s,a) = w \times Q_{model-based}(s,a) + (1-w) \times Q_{model-free}(s,a),$$

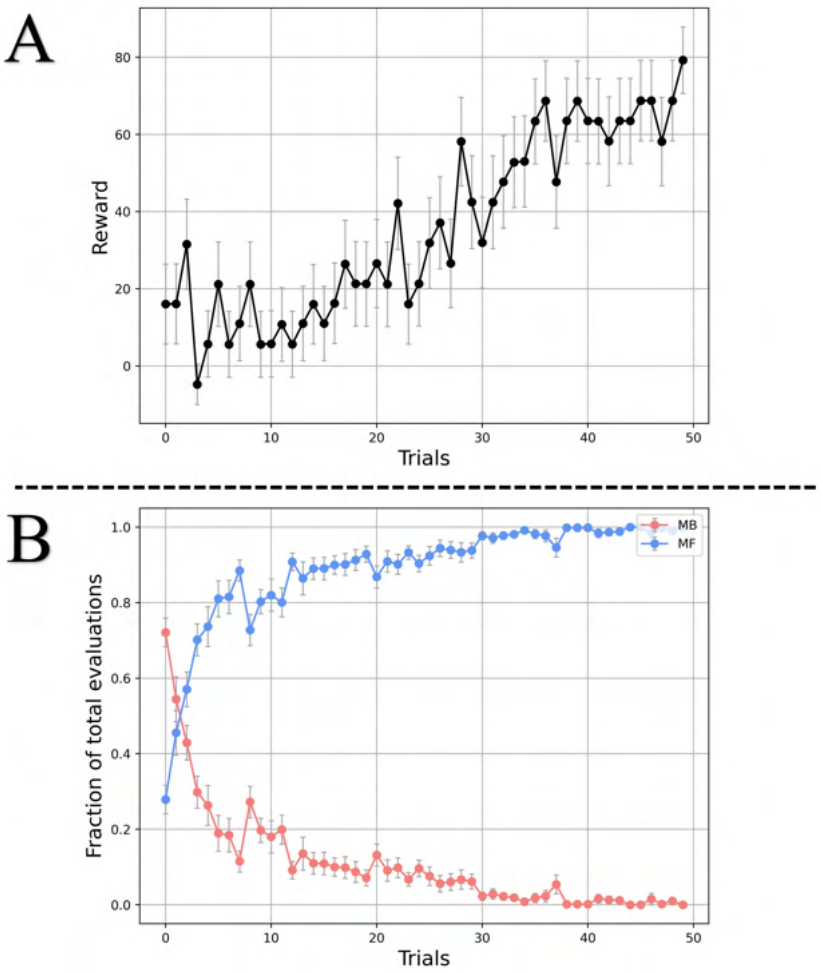

Figure 5.6 (A) A simulation run using the VoI based arbitration. The reward obtained is plotted across trials. The plot is generated by simulating the model for 20 runs. The mean and SEM are plotted. (B) The dual process arbitration across trials. The fraction of total evaluations are plotted for both - MB and MF controller.

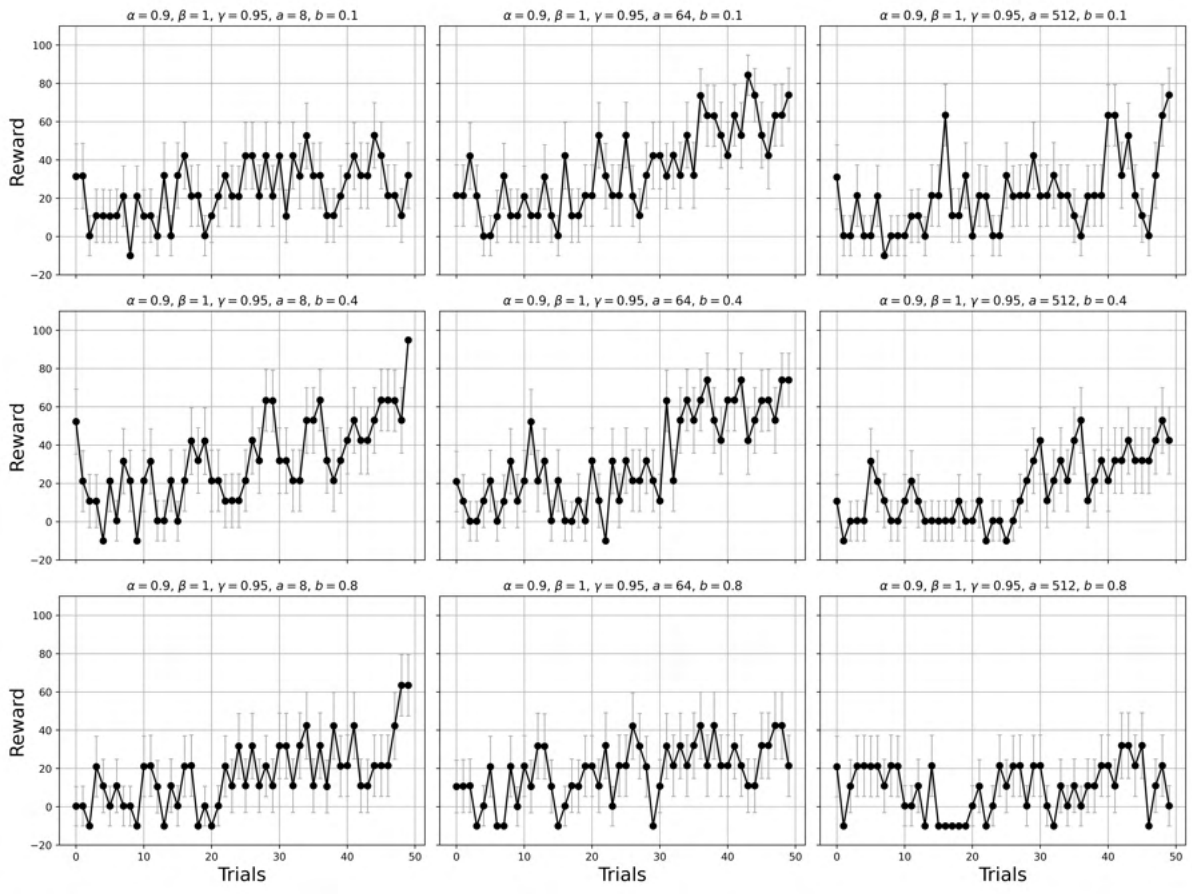

Figure 5.7 Simulations of VoI-based arbitration with different model parameters. The parameters a and b vary across simulations. Each reward plot is generated by simulating the model with the given parameters for 10 runs. The mean and SEM are plotted.

where w is the trial-specific weight term. The weight term w_t for the trial t is computed as

$$w_t = k \times e^{-lt},$$

where k and l are free parameters describing the offset and slope of the function, respectively. The probabilistic action selection is implemented using a softmax function (as described earlier). Figure 5.9 shows a simulation run using the weighted-hybrid arbitration.

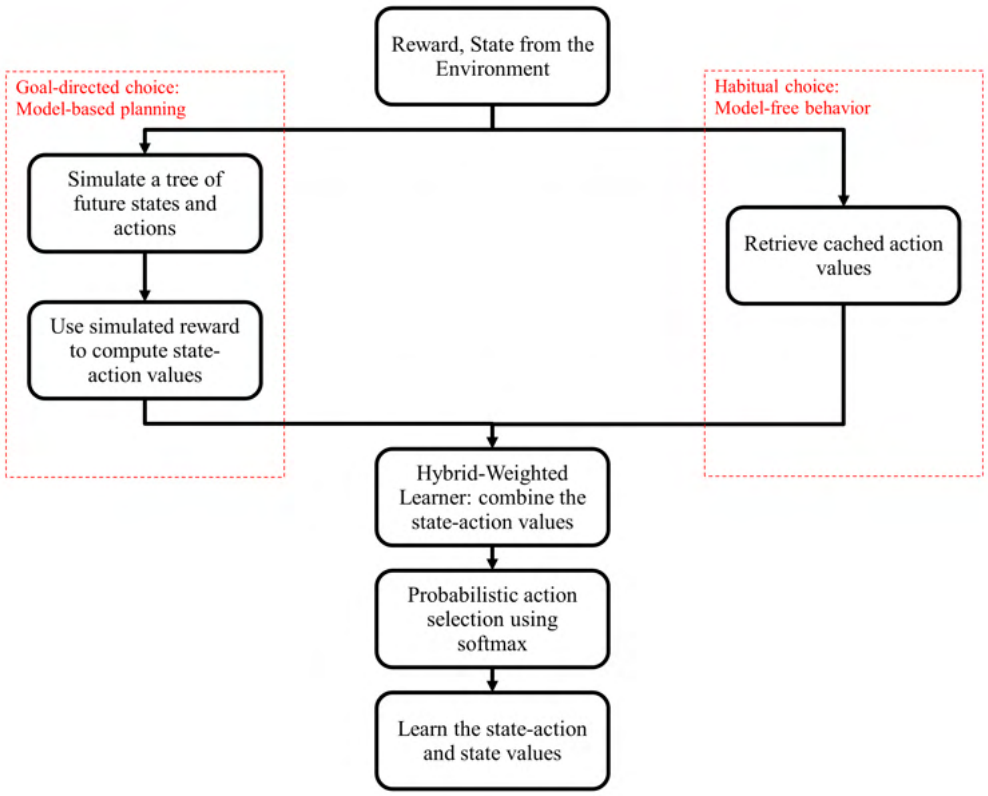

Figure 5.8 A schematic diagram of the weighted-hybrid arbitration (Adapted from Gläscher et al. (2010)).

On simulating the model, we observed that the agent is quickly able to learn the environment. The reward obtained increased with the trials, implying that the agent learned to navigate to the goal position while maximizing the reward score. We also plotted the dual-process arbitration. We observed that the trial-specific weight parameter w decreases over the practice. This suggests that the weight of the model-based estimates decreased over time. In the early training phase, the weighted-hybrid estimates were predominantly governed by the model-based controller. With practice, we observed that the w decreased and $1 - w$ increased. In the late training phase, the weighted-hybrid estimates were largely determined by the model-free controller. Once the agent learned the environment dynamics, it employed this knowledge to figure out the optimal course of action.

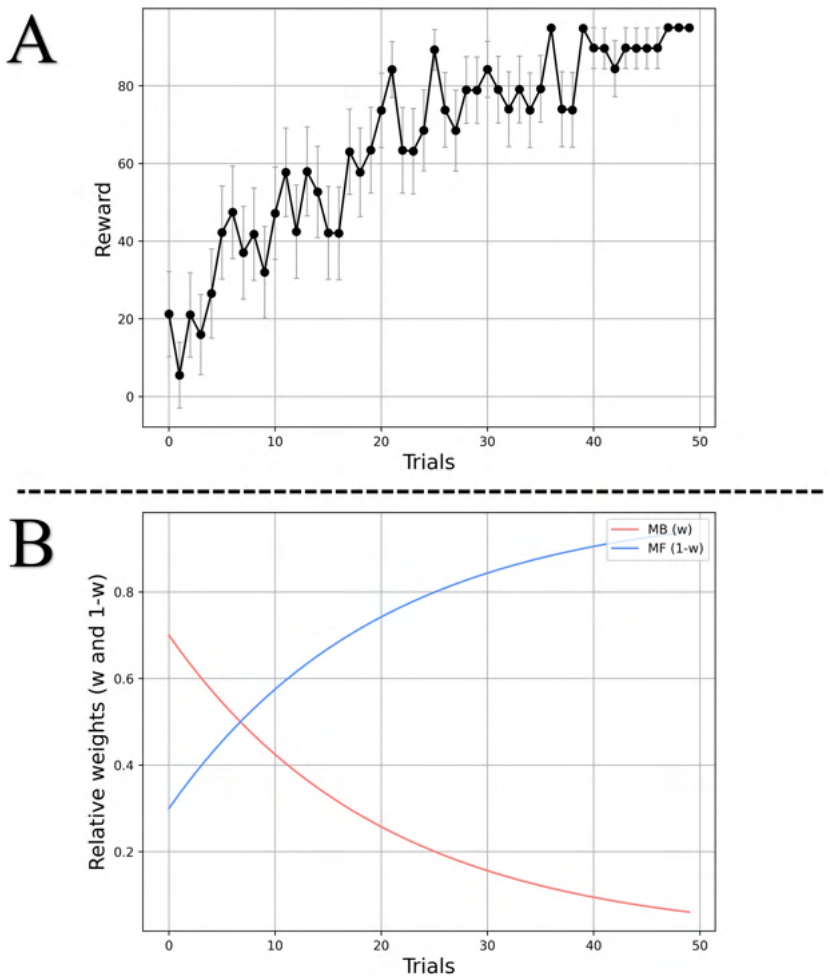

Figure 5.9 (A) A simulation run using the weighted-hybrid arbitration. The reward obtained is plotted across trials. The plot is generated by simulating the model for 20 runs. The mean and SEM are plotted. (B) The dual process arbitration across trials. The relative weights (w and $1 - w$) are plotted for both - MB and MF controller.

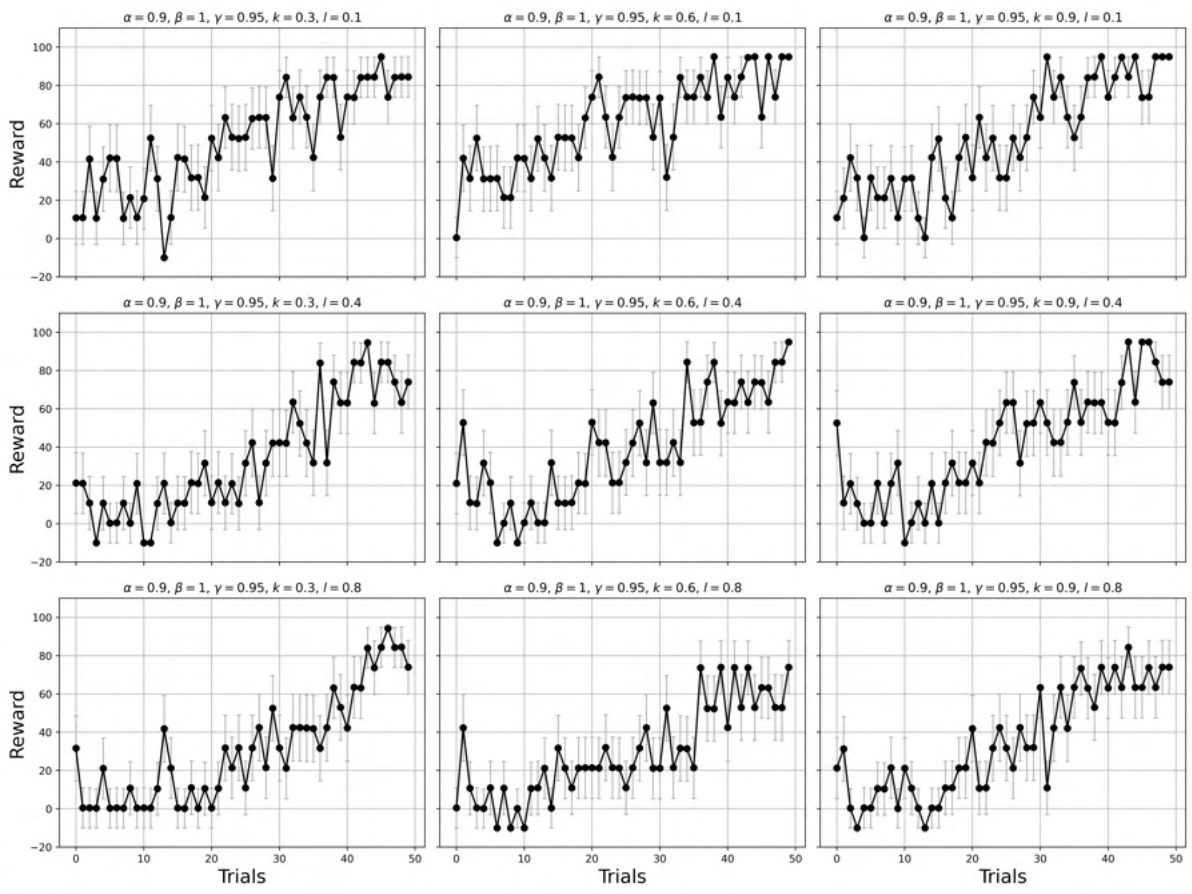

Figure 5.10 Simulations of weighted-hybrid arbitration with different model parameters. The parameters k and l vary across simulations. Each reward plot is generated by simulating the model with the given parameters for 10 runs. The mean and SEM are plotted.

5.4 Simulation

To compare the performance, we simulated VoI based arbitration, weighted-hybrid arbitration, SARSA and random models on the Mixed-SG condition (see Figure 5.11). We observed that VoI based arbitration, weighted-hybrid arbitration and SARSA perform much better than the random baseline. Moreover, VoI based arbitration performs better than SARSA in the early training phase because of the advantage of model-based planning. However, in the late training phase, the performance of SARSA matches that of the VoI based arbitration. We observed that weighted-hybrid arbitration outperforms all the other models. In the initial trials, its performance matches that of VoI based arbitration because both models predominantly rely on model-based planning. In the late phase, weighted-hybrid arbitration returns a greater mean reward as compared to VoI based arbitration.

The simulation experiments assume that the agent has already learned the transition table from the Single-SG condition (see Chapter 2, Experiment-1). We introduced the trial time limit in the environment by capping the maximum number of moves in each trial at 10. A penalty of -1 point was introduced for each move so that the agent learns the optimal trajectory to the goal position.

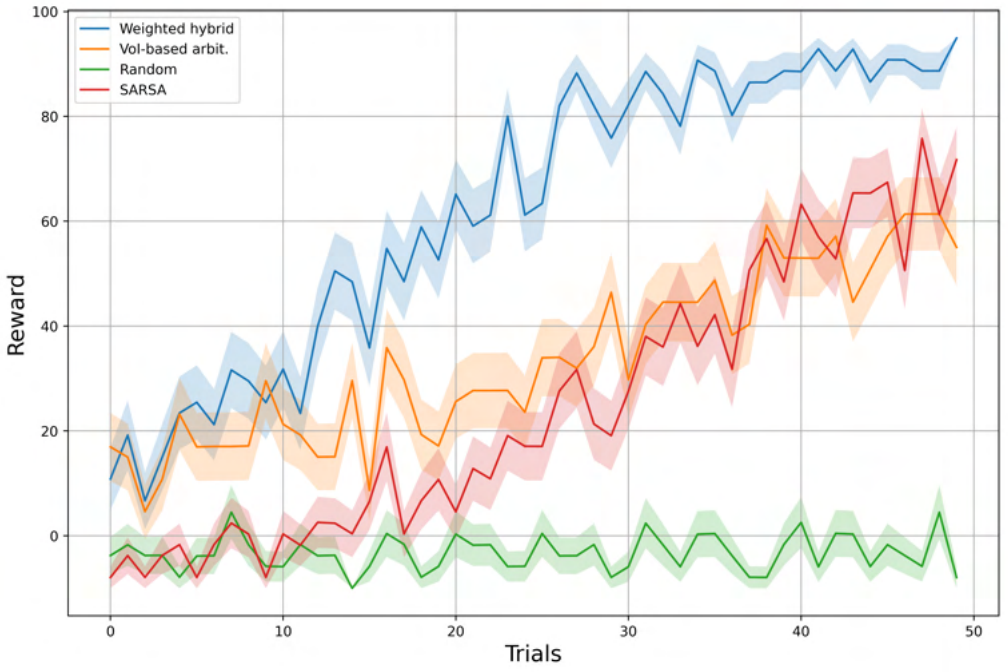

Figure 5.11 Performance comparison. The plot is generated by simulating each model with the given parameters for 50 runs. The mean and SEM are plotted.

5.5 Model-fitting

5.5.1 Experimental Data

The computational models were fit to the data from Chapter 2, Experiment-2 (Mixed-SG condition). Forty-two participants performed Experiment-2 after completing Experiment-1. Experiment-2 involved performing GST on a randomized order of two SG conditions.

In Experiment-1, the participants performed GST on the same SG condition for 41 successful trials (excluding the rotation condition). In Experiment-1, the participants learned the KM by trial-and-error to associate the keypress buttons with the movement directions. In the computational model, the possible movement directions are explicitly known to the RL agent as the action tuple. Experiment-1 can be treated as a free-choice session (without rewards), wherein the transition structure of the environment is learned. Therefore, the model-based RL in our computational model assumes the knowledge of the transition structure (from Experiment-1). Unlike in the actual experiment, the environment penalized each move with -1 so that the agent learns to reach the goal position in a minimum number of steps.

5.5.2 Procedure

In order to assess if the computational model is able to explain the behavioral data, we performed model fitting using the maximum likelihood estimation (MLE). It involves finding the best-fit parameters $\hat{\theta_m}$ for model m. The best-fit parameters are found by maximizing the likelihood $p(d_{1:T}|\theta_m, m)$ of the data $d_{1:T}$, given the model parameters. For reasons of numerical stability, we maximize the log-likelihood

$$LL = \log p(d_{1:T}|\theta_m, m)$$

The log-likelihood can be re-written as

$$LL = \log p(d_{1:T}|\theta_m, m) = \log(\prod_{t=1}^{T} p(a_t|d_{1:t-1}, s_t, \theta_m, m)),$$

where $p(a_t|d_{1:t-1}, s_t, \theta_m, m)$ is the probability of choosing each action a in a given state s_t, given the parameters of the model θ_m. The equation can be written as

$$\log p(d_{1:T}|\theta_m, m) = \sum_{t=1}^{T} \log p(a_t|d_{1:t-1}, s_t, \theta_m, m)$$

The best-fit parameters $\hat{\theta}$ for the model can be found by maximizing the log-likelihood LL or minimizing the negative log-likelihood $NegLL = -LL$. The best-fit parameters were obtained by minimizing the sum of negative log-likelihood of the actions, across all the trials and participants. In line with previous studies (Gershman et al., 2009; Gläscher et al., 2010, 2009), we fitted the computational models to estimate a single set of parameter values across all the participants because it introduces a simple regularization that facilitates stable estimation of the model parameters.

The goodness-of-fit was measured by Bayesian Information Criterion (BIC), which takes into account the maximum likelihood as well as the number of the model parameters.

$$BIC = -2\log(\hat{LL}) + k_m \log(n),$$

where $\hat{LL}$ is the maximum log-likelihood value obtained after model fitting, k is the number of free parameters in model m and n is the total number of observations or data samples.

5.5.3 Results

We performed model-fit using four computational models - Random, SARSA, VoI based arbitration and weighted-hybrid arbitration. The model-fitting procedure was performed using the `minimize` function from scipy.optimize package (Virtanen et al., 2020). We performed the model fitting procedure for each model 10 times while sampling the initial parameters from a uniform distribution at the start of every run. The mean negative log-likelihood and BIC values are reported. We found that the weighted-hybrid arbitration model explained the behavioral data better than other models (see Table 5.1).

(A)

Model	**Neg LL**	**BIC**
Random	18342.96	36685.91
SARSA	12802.81	25634.24
Weighted-hybrid arbit.	7074.122	14195.94
VoI based arbit.	8200.875	16449.45

(B)

Parameter	**Value**
Learning rate (α)	0.73
Inv. temp. parameter (β)	0.11
Discount factor (γ)	0.94
Offset parameter (k)	0.75
Slope parameter (l)	0.01

Table 5.1 (A) Model-fit results using different models on the behavioral data. The negative log-likelihood (Neg LL) and Bayesian Information Criterion (BIC) values are reported. Lower BIC values indicate a better fit. (B) Best-fit parameters for the weighted-hybrid arbitration.

5.6 Discussion

The hybrid RL arbitration combines the learning from model-based and model-free controllers to give rise to the behaviors that lie on the spectrum of habitual and goal-directed behaviors. It provides a normative computational framework that can be used to model the role of dual learning processes in

sequence learning. We propose and compare the two hybrid-RL frameworks to provide a computational account of learning in internally-guided sequence learning. The proposed VoI arbitration introduces a cost-benefit analysis for the meta-choice. Model-based planning is only engaged if the benefits of acquiring new knowledge of the environment outweigh the time-varying costs of cognitive efforts required. The simulations show that VoI computation denotes the need for accurately estimating the existing state-action values. For instance, the VoI is high if, in a given state, there is no clear best action. The need for model-based planning subsides as the VoI decreases with substantial experience and knowledge of the environment. In case the rewards are devalued, the arbitration mechanism can engage model-based planning to adapt to the change in the environment. The weighted-hybrid arbitration combines the estimates from the model-based and model-free learner. The values are integrated using a time-varying parameter which denotes the relative weights of each controller. We observed that weighted-hybrid arbitration performs better in simulation than other models (see Figure 5.11). This can be attributed to the greater involvement of model-based planning (See Appendix [6.2] for arbitration plots). The learned state transition table and keymap are employed when planning trajectory sequences from the start to the goal position. Therefore, the agent is able to learn the optimal trajectories to the goal position quickly. We also observed that the weighted-hybrid model describes the empirical data better than other models (see Table 5.1).

The skill learning has been explained in the sequential phase-wise acquisition of motor behaviors (Ackerman, 1988; Fitts and Posner, 1967; Haibach et al., 2018). For example, Fitts and Posner (1967) theorize a three-stage account of skill learning. The three sequential stages are cognitive, associative and autonomous. The initial phase involves the dominant role of cognitive learning processes. The cognitive learning phase involves greater demands on the working memory and attentional resources. In internally-guided sequencing, the initial phase would involve the acquisition of the KM, which can be further employed to plan the optimal sequence of movements. Such planning would be akin to model-based learning, which is sample efficient and adaptive but incurs substantial computational costs. The intermediate associative phase is characterized by a gradual decrease in reliance on the cognitive processes. Concurrently, the role of motor processes steadily increase. The associative phase involves a switch from the dominant role of model-based planning to the model-free behavior. The late autonomous phase is dominated by the motor learning processes. With the acquisition of the environment model, the agent is certain about the consequences of its actions and, therefore, relies on the inexpensive model-free behavior. The simulation results show that the dual-process arbitration matches Fitts' three phases account of learning. In the VoI based arbitration (see Figure 5.6 B) and the weighted-hybrid arbitration (see Figure 5.9 B), we observe that the model-based processor is dominant in the initial cognitive phase. In the late motor phase, we observe a dominant role of the model-free controller as the reliance on model-based planning declines. The computational account of arbitration indicates a transition from goal-directed to habitual learning. The cognitive phase of skill learning is characterized by goal-directed planning, whereas the motor phase involves habitual learning. This is in line with the find-

ings from our inter-manual transfer task (see Chapter 4), where we showed that the learning transition from effector-independent cognitive learning to effector-dependent motor learning with practice.

Moreover, our results confirm the findings from Fermin et al. (2010) in internally-guided sequencing. In line with their account, once the action space is explored, the early learning phase involves a dominant role of the model-based controller, whereas the late phase involves a dominant role of the model-free controller. We show that each controller takes a dominant role in the stage-wise transitions in learning based on the extent of practice and the knowledge of the environment.

Chapter 6

Conclusion and Future Directions

6.1 Summary of the Main Findings

We outline the premise of our investigation in the fundamental difference between externally-specified sequencing and internally-guided sequencing. We argue that sequential finger movements while performing grid-navigation tasks amounts to learning internally-guided sequence. Using Grid-Sailing Task as our canonical paradigm, we present an empirical and computational investigation of skill learning in internally-guided sequencing.

In the first study, we aimed to investigate the learning processes involved in internally-guided sequencing. We provide evidence for cognitive and motor learning in internally-guided sequencing. We show increasing dexterity on repeated execution of the same trajectories as evidence for motor learning. Using a visuomotor rotation, we show that the performance improvements cannot be solely attributed to general motor improvements. We further hypothesize the role of cognitive learning in GST. The role of such a cognitive learning process is confirmed by showing transfer-related performance improvements on the randomized and mixed order of previously unseen SG conditions. We show that the participants were able to generalize and transfer a KM-specific internal model that facilitated performance.

In the second study, we provide evidence for the practice-driven performance improvements in GST due to motor chunking. In contrast with previous studies, we show chunking in a more realistic and practical motor paradigm which is internally-guided. Our findings show evidence for spontaneous chunking without pre-specified or externally-guided structures while replicating the earlier results with a less constrained experimental paradigm. We show how the chunks evolve and re-organize during various phases of the practice.

In the third study, we employ an inter-manual transfer task to investigate the stage-wise progression of learning in GST. We show that the early practice phase involves the dominant role of the effector-independent cognitive learning process, whereas the late practice phase involves the dominant role of the effector-dependent motor learning process. We confirm the cognitive-to-motor transition of learning in internally-guided sequencing.

In the last chapter, we propose a computational account of skill learning in internally-guided sequencing. We situate the typical behavioral phenomenon in GST in the dual-process account of skill learning and discuss the computational analogues of goal-directed and habitual learning processes. We proposed two reinforcement learning schemes that describe the arbitration between the goal-directed model-based controller and the habitual model-free controller. Using simulations and computational model fitting, we show that the hybrid-weighted model describes the behavioral data better than other models.

6.2 Limitations and Future Work

Future work can attempt to investigate the nature of cognitive learning in internally-guided sequencing. For instance, the experiment in Chapter 2 can be modified to statistically examine the differences in trajectory traversals in different conditions. A hypothetical measure of trajectory density can be used to understand the evolution of trajectories as the sequence is learned. A categorical comparison of the trajectory features can reveal further evidence for the "transfer" of KM-specific learning. Moreover, the role of KM-specific learning can be further validated by introducing a new KM on the same SG conditions. Future studies can also employ a retention task by extending the experimental task over a period of days to dissociate the cognitive and motor learning in GST. We anticipate that the KM-specific internal model will be retained longer than the fine-tuned motor movements specific to the trajectory. Consequently, we can expect to see that the participants quickly recall the learned KM and perform better than they had initially performed at the beginning of the task.

Previous studies (Bapi et al., 2005; Rosenbaum et al., 1983) have shown that chunking occurs as a hierarchical organization of motor action sequences in externally-guided visuomotor sequencing. Future work in internally-guided sequencing can experimentally probe theoretical questions about hierarchical re-organization—i.e., how chunks are formed and integrated into a multi-level structure. Subsequent studies can also benefit from other methods of identifying and segmenting chunks in internally guided sequencing. For example, Acuna et al. (2014) proposes a Bayesian algorithm that identifies chunks based on response times, errors and their correlations. Wymbs et al. (2012) used a multi-trial community detection approach after constructing the sequence network.

While reinforcement learning is used to model the choice data in sequencing tasks, the response time data is largely overlooked in the existing models. Sequential sampling models provide a normative way of formalizing the decision processes (Ratcliff et al., 2016). The reinforcement learning and sequential sampling models can be integrated into a unified framework for the joint modeling of choice and response time data (Miletić et al., 2020). Such models would prove very useful in understanding the trial-wise dynamics of choice decisions and response times underlying sequencing in internally-guided paradigms.

Related Publications

Journal Publications

1. Bera, K., Shukla, A., & Bapi, R. S. (2021). Cognitive and Motor Learning in Internally-Guided Motor Skills. Frontiers in Psychology, 12, 1035.

2. Bera, K., Shukla, A., & Bapi, R. S. (2021). Motor chunking in internally guided sequencing. Brain Sciences, 11(3), 292.

Conference Publications

1. Bera, K., Mandilwar, Y., & Raju, B. (2019). Value-of-Information based Arbitration between Model-based and Model-free Control (Oral). Sixth Annual Conference of the Cognitive Science Society, Goa, IN. arXiv:1912.05453.

2. Bera, K., Shukla, A., & Raju, B. (2021). Motor chunking in Internally-guided sequence learning (Abstract). Seventh Annual Conference of the Cognitive Science Society, Virtual meeting.

3. Bera, K., Shukla, A., & Raju, B. (2020). Motor Chunking During Sequence Learning in Grid-Navigation Tasks (Abstract). In Proceedings of the 42nd Annual Meeting of the Cognitive Science Society. Toronto, CA.

4. Bera, K., Shukla, A., & Raju, B. (2020). Grid-navigation tasks involve skill learning (Abstract). In Proceedings of the 42nd Annual Meeting of the Cognitive Science Society. Toronto, CA.

5. Bera, K., Mandilwar, Y., Shukla, A., & Raju, B. (2020). Value-of-Information based Arbitration between Model-based and Model-free Control (Abstract). In Proceedings of the 42nd Annual Meeting of the Cognitive Science Society. Toronto, CA.

6. Bera, K., Savalia, T., & Raju, B. (2018). A Computational Framework for Motor Skill Acquisition (Poster). Fifth Annual Conference of the Cognitive Science Society, Guwahati, IN. arXiv preprint arXiv:1901.01856.

Bibliography

Abrahamse, E. L., Jiménez, L., Verwey, W. B., and Clegg, B. A. (2010). Representing serial action and perception. *Psychonomic Bulletin & Review*, 17(5):603–623.

Abrahamse, E. L., Ruitenberg, M. F. L., de Kleine, E., and Verwey, W. B. (2013). Control of automated behavior: insights from the discrete sequence production task. *Frontiers in Human Neuroscience*, 7.

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, 117(3):288–318. Place: US Publisher: American Psychological Association.

Acuna, D. E., Wymbs, N. F., Reynolds, C. A., Picard, N., Turner, R. S., Strick, P. L., Grafton, S. T., and Kording, K. P. (2014). Multifaceted aspects of chunking enable robust algorithms. *Journal of Neurophysiology*, 112(8):1849–1856.

Adams, C. D. and Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *The Quarterly Journal of Experimental Psychology Section B*, 33(2):109–121. Publisher: Routledge _eprint: https://doi.org/10.1080/14640748108400816.

Balleine, B. W. and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology*, 37(4):407–419.

Balleine, B. W. and O'Doherty, J. P. (2010). Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, 35(1):48–69.

Bapi, R. S., Doya, K., and 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(2):149–162.

Bapi, R. S., Miyapuram, K., Graydon, F., and Doya, K. (2006). fMRI investigation of cortical and subcortical networks in the learning of abstract and effector-specific representations of motor sequences. *NeuroImage*, 32(2):714–727.

Bapi, R. S., Pammi, V. S. C., Miyapuram, K. P., and Ahmed (2005). Investigation of sequence processing: A cognitive and computational neuroscience perspective. *Current Science*, 89(10):1690–1698. Publisher: Current Science Association.

Barnhoorn, J. S., Van Asseldonk, E. H. F., and Verwey, W. B. (2019). Differences in chunking behavior between young and older adults diminish with extended practice. *Psychological Research*, 83(2):275–285.

Bera, K., Shukla, A., and Bapi, R. S. (2021a). Cognitive and Motor Learning in Internally-Guided Motor Skills. *Frontiers in Psychology*, 12:604323.

Bera, K., Shukla, A., and Bapi, R. S. (2021b). Motor Chunking in Internally Guided Sequencing. *Brain Sciences*, 11(3):292.

Bo, J. and Seidler, R. D. (2009). Visuospatial Working Memory Capacity Predicts the Organization of Acquired Explicit Motor Sequences. *Journal of Neurophysiology*, 101(6):3116–3125.

Boutin, A., Badets, A., Salesse, R. N., Fries, U., Panzer, S., and Blandin, Y. (2012). Practice makes transfer of motor skills imperfect. *Psychological Research*, 76(5):611–625.

Chen, X., Mohr, K., and Galea, J. M. (2017). Predicting explorative motor learning using decision-making and motor noise. *PLOS Computational Biology*, 13(4):e1005503. Publisher: Public Library of Science.

Clegg, B. A., DiGirolamo, G. J., and Keele, S. W. (1998). Sequence learning. *Trends in Cognitive Sciences*, 2(8):275–281. Publisher: Elsevier.

Curran, T. and Keele, S. W. (1993). Attentional and nonattentional forms of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(1):189–202.

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12):1704–1711. Number: 12 Publisher: Nature Publishing Group.

Dayan, P. (2009). Goal-directed control and its antipodes. *Neural Networks: The Official Journal of the International Neural Network Society*, 22(3):213–219.

Dezfouli, A. and Balleine, B. W. (2012). Habits, action sequences and reinforcement learning: Habits and action sequences. *European Journal of Neuroscience*, 35(7):1036–1051.

Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Networks*, 12(7-8):961–974.

Doya, K. (2000). Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current Opinion in Neurobiology*, 10(6):732–739.

Drucker, J. H., Sathian, K., Crosson, B., Krishnamurthy, V., McGregor, K. M., Bozzorg, A., Gopinath, K., Krishnamurthy, L. C., Wolf, S. L., Hart, A. R., Evatt, M., Corcos, D. M., and Hackney, M. E. (2019). Internally Guided Lower Limb Movement Recruits Compensatory Cerebellar Activity in People With Parkinson's Disease. *Frontiers in Neurology*, 10:537.

Fermin, A., Yoshida, T., Ito, M., Yoshimoto, J., and Doya, K. (2010). Evidence for Model-Based Action Planning in a Sequential Finger Movement Task. *Journal of Motor Behavior*, 42(6):371–379.

Fermin, A. S. R., Yoshida, T., Yoshimoto, J., Ito, M., Tanaka, S. C., and Doya, K. (2016). Model-based action planning involves cortico-cerebellar and basal ganglia networks. *Scientific Reports*, 6(1):31378.

Fitts, P. M. and Posner, M. I. (1967). *Human performance*. Brooks/Cole Pub. Co., Belmont, Calif.

Galea, J. M., Vazquez, A., Pasricha, N., de Xivry, J.-J. O., and Celnik, P. (2011). Dissociating the roles of the cerebellum and motor cortex during adaptive learning: the motor cortex retains what the cerebellum learns. *Cerebral Cortex (New York, N.Y.: 1991)*, 21(8):1761–1770.

Gershman, S. J., Pesaran, B., and Daw, N. D. (2009). Human Reinforcement Learning Subdivides Structured Action Spaces by Learning Effector-Specific Values. *The Journal of Neuroscience*, 29(43):13524–13531.

Ghilardi, M. F., Moisello, C., Silvestri, G., Ghez, C., and Krakauer, J. W. (2009). Learning of a Sequential Motor Skill Comprises Explicit and Implicit Components That Consolidate Differently. *Journal of Neurophysiology*, 101(5):2218–2229.

Gläscher, J., Daw, N., Dayan, P., and O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, 66(4):585–595.

Gläscher, J., Hampton, A. N., and 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 (New York, NY)*, 19(2):483–495.

Gowen, E. and Miall, R. (2007). Differentiation between external and internal cuing: An fMRI study comparing tracing with drawing. *NeuroImage*, 36(2):396–410.

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annual Review of Neuroscience*, 31:359–387.

Greenwald, A. G. (1970). Sensory feedback mechanisms in performance control: With special reference to the ideo-motor mechanism. *Psychological Review*, 77(2):73–99.

Haibach, P. S., Reid, G., and Collier, D. H. (2018). *Motor learning and development*. Human Kinetics, Champaign, IL, second edition edition.

Haith, A. M. and Krakauer, J. W. (2013). Model-Based and Model-Free Mechanisms of Human Motor Learning. *Advances in experimental medicine and biology*, 782:1–21.

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. *Brain mechanisms and learning*, 37:46. Publisher: Oxford: Blackwell.

Herwig, A. and Waszak, F. (2009). Intention and attention in ideomotor learning. *Quarterly Journal of Experimental Psychology*, 62(2):219–227.

Hicks, R. E. (1974). Asymmetry of Bilateral Transfer. *The American Journal of Psychology*, 87(4):667–674. Publisher: University of Illinois Press.

Hikosaka, O., Nakahara, H., Rand, M. K., Sakai, K., Lu, X., Nakamura, K., Miyachi, S., and Doya, K. (1999). Parallel neural networks for learning sequential procedures. *Trends in Neurosciences*, 22(10):464–471.

Hikosaka, O., Nakamura, K., Sakai, K., and Nakahara, H. (2002). Central mechanisms of motor skill learning. *Current Opinion in Neurobiology*, 12(2):217–222.

Hikosaka, O., Rand, M. K., Miyachi, S., and Miyashita, K. (1995). Learning of sequential movements in the monkey: process of learning and retention of memory. *Journal of Neurophysiology*, 74(4):1652–1661.

Hoffmann, J., Sebald, A., and Stöcker, C. (2001). Irrelevant response effects improve serial learning in serial reaction time tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(2):470–482.

Hommel, B. (2003). Acquisition and control of voluntary action. In *Voluntary action: Brains, minds, and sociality.*, pages 34–48. Oxford University Press, New York, NY, US.

Hommel, B., Müsseler, J., Aschersleben, G., and Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24(5):849–878.

Huang, V. S., Haith, A., Mazzoni, P., and Krakauer, J. W. (2011). Rethinking motor learning and savings in adaptation paradigms: model-free memory for successful actions combines with internal models. *Neuron*, 70(4):787–801.

Izawa, J., Rane, T., Donchin, O., and Shadmehr, R. (2008). Motor Adaptation as a Process of Reoptimization. *Journal of Neuroscience*, 28(11):2883–2891.

JASP Team (2020). JASP (Version 0.14.1)[Computer software].

Jueptner, J., Jueptner, M., Jenkins, I., Brooks, D., Frackowiak, R., and Passingham, R. (1996). The sensory guidance of movement: a comparison of the cerebellum and basal ganglia. *Experimental Brain Research*, 112(3).

Jueptner, M. (1998). A review of differences between basal ganglia and cerebellar control of movements as revealed by functional imaging studies. *Brain*, 121(8):1437–1449.

Keele, S. W., Ivry, R., Mayr, U., Hazeltine, E., and Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychological Review*, 110(2):316–339.

Keele, S. W., Jennings, P., Jones, S., Caulton, D., and Cohen, A. (1995). On the Modularity of Sequence Representation. *Journal of Motor Behavior*, 27(1):17–30.

Kennerley, S. W., Sakai, K., and Rushworth, M. (2004). Organization of Action Sequences and the Role of the Pre-SMA. *Journal of Neurophysiology*, 91(2):978–993.

Keramati, M., Smittenaar, P., Dolan, R. J., and Dayan, P. (2016). Adaptive integration of habits into depth-limited planning defines a habitual-goal–directed spectrum. *Proceedings of the National Academy of Sciences*, 113(45):12868–12873. Publisher: National Academy of Sciences Section: Biological Sciences.

Khamassi, M. and Humphries, M. D. (2012). Integrating cortico-limbic-basal ganglia architectures for learning model-based and model-free navigation strategies. *Frontiers in Behavioral Neuroscience*, 6.

Koch, I. and Hoffmann, J. (2000). Patterns, chunks, and hierarchies in serial reaction-time tasks. *Psychological Research Psychologische Forschung*, 63(1):22–35.

Kovacs, A. J., Mühlbauer, T., and Shea, C. H. (2009). The coding and effector transfer of movement sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2):390–407.

Kumar, A., Panthi, G., Divakar, R., and Mutha, P. K. (2020). Mechanistic determinants of effector-independent motor memory encoding. *Proceedings of the National Academy of Sciences*, 117(29):17338–17347.

Lashley, K. S. (1951). The problem of serial order in behavior. In *Cerebral mechanisms in behavior; the Hixon Symposium.*, pages 112–146. Wiley, Hoboken, NJ (USA).

Lee, J.-Y. and Schweighofer, N. (2009). Dual Adaptation Supports a Parallel Architecture of Motor Memory. *Journal of Neuroscience*, 29(33):10396–10404. Publisher: Society for Neuroscience Section: Articles.

Lee, S. W., Shimojo, S., and O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron*, 81(3):687–699.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95(4):492–527.

Logan, G. D. (2018). Automatic control: How experts act without thinking. *Psychological Review*, 125(4):453–485.

McDougle, S. D. and Taylor, J. A. (2019). Dissociable cognitive strategies for sensorimotor learning. *Nature Communications*, 10(1):40.

Miletić, S., Boag, R. J., and Forstmann, B. U. (2020). Mutual benefits: Combining reinforcement learning with sequential sampling models. *Neuropsychologia*, 136:107261.

Mushiake, H., Saito, N., Sakamoto, K., Itoyama, Y., and Tanji, J. (2006). Activity in the Lateral Prefrontal Cortex Reflects Multiple Steps of Future Events in Action Plans. *Neuron*, 50(4):631–641.

Nattkemper, D., Ziessler, M., and Frensch, P. A. (2010). Binding in voluntary action control. *Neuroscience & Biobehavioral Reviews*, 34(7):1092–1101.

Newell, K. M. (1991). Motor Skill Acquisition. *Annual Review of Psychology*, 42(1):213–237. Publisher: Annual Reviews.

Nissen, M. J. and Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, 19(1):1–32.

O'Doherty, J. P. (2011). Contributions of the ventromedial prefrontal cortex to goal-directed action selection. *Annals of the New York Academy of Sciences*, 1239:118–129.

Panzer, S., Krueger, M., Muehlbauer, T., Kovacs, A. J., and Shea, C. H. (2009). Inter-manual transfer and practice: Coding of simple motor sequences. *Acta Psychologica*, 131(2):99–109.

Park, J.-H. and Shea, C. H. (2003). Effect of Practice on Effector Independence. *Journal of Motor Behavior*, 35(1):33–40.

Parlow, S. E. and Kinsbourne, M. (1990). Asymmetrical transfer of braille acquisition between hands. *Brain and Language*, 39(2):319–330.

Penhune, V. B. and Steele, C. J. (2012). Parallel contributions of cerebellar, striatal and M1 mechanisms to motor sequence learning. *Behavioural Brain Research*, 226(2):579–591.

Perez, M., Tanaka, S., Wise, S., Sadato, N., Tanabe, H., Willingham, D., and Cohen, L. (2007). Neural Substrates of Intermanual Transfer of a Newly Acquired Motor Skill. *Current Biology*, 17(21):1896–1902.

Pezzulo, G., Rigoli, F., and Chersi, F. (2013). The mixed instrumental controller: using value of information to combine habitual choice and mental simulation. *Frontiers in Psychology*, 4:92.

Povel, D.-J. and Collard, R. (1982). Structural factors in patterned finger tapping. *Acta Psychologica*, 52(1-2):107–123.

Prinz, W. (1997). Perception and Action Planning. *European Journal of Cognitive Psychology*, 9(2):129–154.

Ramkumar, P., Acuna, D. E., Berniker, M., Grafton, S. T., Turner, R. S., and Kording, K. P. (2016). Chunking as the result of an efficiency computation trade-off. *Nature Communications*, 7(1):12176.

Ratcliff, R., Smith, P. L., Brown, S. D., and McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, 20(4):260–281.

Restle, F. and Burnside, B. L. (1972). Tracking of serial patterns. *Journal of Experimental Psychology*, 95(2):299–307.

Robertson, E. M. (2007). The Serial Reaction Time Task: Implicit Motor Skill Learning? *Journal of Neuroscience*, 27(38):10073–10075.

Rosenbaum, D. A., Kenny, S. B., and Derr, M. A. (1983). Hierarchical control of rapid movement sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 9(1):86–102.

Sakai, K., Kitaguchi, K., and Hikosaka, O. (2003). Chunking during human visuomotor sequence learning. *Experimental Brain Research*, 152(2):229–242.

Savalia, T., Shukla, A., and Bapi, R. S. (2016). A Unified Theoretical Framework for Cognitive Sequencing. *Frontiers in Psychology*, 7.

Schmidt, R. A., Lee, T. D., Winstein, C. J., Wulf, G., and Zelaznik, H. N. (2019). *Motor control and learning: a behavioral emphasis*. Human Kinetics, Champaign, IL, sixth edition edition.

Shadmehr, R. and Krakauer, J. W. (2008). A computational neuroanatomy for motor control. *Experimental Brain Research*, 185(3):359–381.

Shadmehr, R. and Mussa-Ivaldi, F. (1994). Adaptive representation of dynamics during learning of a motor task. *The Journal of Neuroscience*, 14(5):3208–3224.

Shea, C. H., Kovacs, A. J., and Panzer, S. (2011). The Coding and Inter-Manual Transfer of Movement Sequences. *Frontiers in Psychology*, 2. Publisher: Frontiers.

Smith, M. A., Ghazizadeh, A., and Shadmehr, R. (2006). Interacting Adaptive Processes with Different Timescales Underlie Short-Term Motor Learning. *PLOS Biology*, 4(6):e179. Publisher: Public Library of Science.

Stadler, M. A. (1993). Implicit serial learning: Questions inspired by Hebb (1961). *Memory & Cognition*, 21(6):819–827.

Sutton, R. S. (1991). Planning by Incremental Dynamic Programming. In Birnbaum, L. A. and Collins, G. C., editors, *Machine Learning Proceedings 1991*, pages 353–357. Morgan Kaufmann, San Francisco (CA).

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: an introduction*. MIT Press, Cambridge, MA, second edition. OCLC: 1190775239.

Taylor, H. G. and Heilman, K. M. (1980). Left-hemisphere motor dominance in righthanders. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 16(4):587–603.

Thalmann, M., Souza, A. S., and Oberauer, K. (2019). How does chunking help working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(1):37–55.

Thomas, T., Miyapuram, K. P., and Bapi, R. S. (2012). Inter-manual transfer of visuo-motor sequence learning. In *Proceedings of the XXII Annual Convention of the National Academy of Psychology (NAOP)*, Bangalore, IN.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, 55(4):189–208. Place: US Publisher: American Psychological Association.

van Donkelaar, P., Stein, J. F., Passingham, R. E., and Miall, R. C. (1999). Neuronal Activity in the Primate Motor Thalamus During Visually Triggered and Internally Generated Limb Movements. *Journal of Neurophysiology*, 82(2):934–945.

van Mier, H. I. and Petersen, S. E. (2006). Intermanual transfer effects in sequential tactuomotor learning: Evidence for effector independent coding. *Neuropsychologia*, 44(6):939–949.

Verwey, W. B. (1996). Buffer loading and chunking in sequential keypressing. *Journal of Experimental Psychology: Human Perception and Performance*, 22(3):544–562.

Verwey, W. B. (2001). Concatenating familiar movement sequences: the versatile cognitive processor. *Acta Psychologica*, 106(1-2):69–95.

Verwey, W. B., Abrahamse, E. L., and Jiménez, L. (2009). Segmentation of short keying sequences does not spontaneously transfer to other sequences. *Human Movement Science*, 28(3):348–361.

Verwey, W. B. and Clegg, B. A. (2005). Effector dependent sequence learning in the serial RT task. *Psychological Research Psychologische Forschung*, 69(4):242–251.

Verwey, W. B. and Dronkert, Y. (1996). Practicing a Structured Continuous Key-Pressing Task: Motor Chunking or Rhythm Consolidation? *Journal of Motor Behavior*, 28(1):71–79.

Verwey, W. B. and Eikelboom, T. (2003). Evidence for Lasting Sequence Segmentation in the Discrete Sequence-Production Task. *Journal of Motor Behavior*, 35(2):171–181.

Verwey, W. B., Shea, C. H., and Wright, D. L. (2015). A cognitive framework for explaining serial processing and sequence execution strategies. *Psychonomic Bulletin & Review*, 22(1):54–77.

Verwey, W. B. and Wright, D. L. (2004). Effector-independent and effector-dependent learning in the discrete sequence production task. *Psychological Research*, 68(1):64–70.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, I., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., and van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272.

Watkins, C. J. C. H. and Dayan, P. (1992). Q-learning. *Machine Learning*, 8(3):279–292.

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

Willingham, D. B. (1999). Implicit motor sequence learning is not purely perceptual. *Memory & Cognition*, 27(3):561–572.

Willingham, D. B., Wells, L. A., Farrell, J. M., and Stemwedel, M. E. (2000). Implicit motor sequence learning is represented in response locations. *Memory & Cognition*, 28(3):366–375.

Wolpert, D. M., Diedrichsen, J., and Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience*, 12(12):739–751. Number: 12 Publisher: Nature Publishing Group.

Wolpert, D. M. and Landy, M. S. (2012). Motor control is decision-making. *Current Opinion in Neurobiology*, 22(6):996–1003.

Wymbs, N., Bassett, D., Mucha, P., Porter, M., and Grafton, S. (2012). Differential Recruitment of the Sensorimotor Putamen and Frontoparietal Cortex during Motor Chunking in Humans. *Neuron*, 74(5):936–946.

Yamaguchi, M. and Logan, G. D. (2014). Pushing typists back on the learning curve: Revealing chunking in skilled typewriting. *Journal of Experimental Psychology: Human Perception and Performance*, 40(2):592–612.

Yin, H. H. and Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, 7(6):464–476. Number: 6 Publisher: Nature Publishing Group.

Yin, H. H., Knowlton, B. J., and Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *The European Journal of Neuroscience*, 19(1):181–189.

Zeigler, H. P. and Gallistel, C. R. (1981). The Organization of Action: A New Synthesis. *The American Journal of Psychology*, 94(1):190.

Ziessler, M. and Nattkemper, D. (2001). Learning of event sequences is based on response-effect learning: Further evidence from a serial reaction task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(3):595–613.

Ziessler, M., Nattkemper, D., and Frensch, P. A. (2004). The role of anticipation and intention in the learning of effects of self-performed actions. *Psychological Research*, 68(2-3):163–175.

Appendix

Chapter 5: Additional Results

Model-fitting

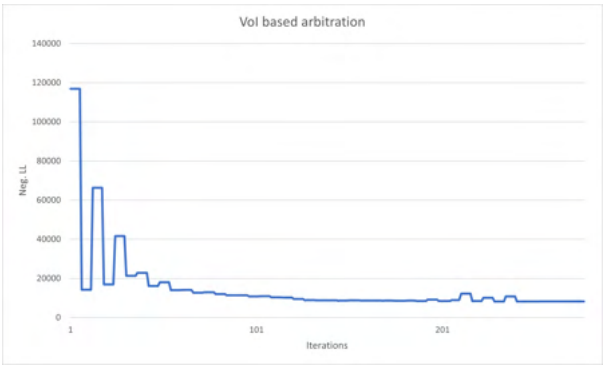

Figure 6.1 Finding the best-fit parameters: The Neg. LL for VoI based arbitration is plotted across optimizer iterations. The mean of all the 10 model-fit runs is plotted.

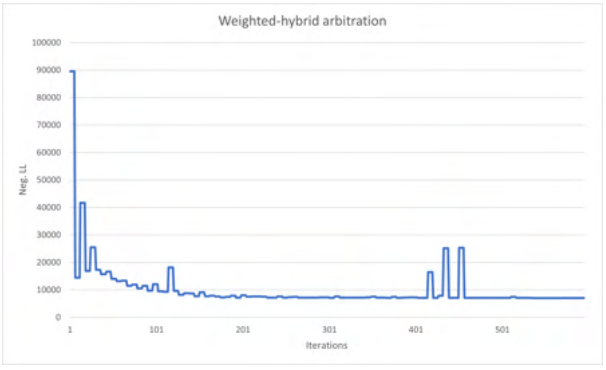

Figure 6.2 Finding the best-fit parameters: The Neg. LL for weighted-hybrid arbitration is plotted across optimizer iterations. The mean of all the 10 model-fit runs is plotted.

Arbitration between Model-based and Model-free Reinforcement Learning

Using the best-fit parameter identified after the model-fitting procedure, we simulated VoI based arbitration and weighted-hybrid arbitration. The arbitration plots for both are plotted.

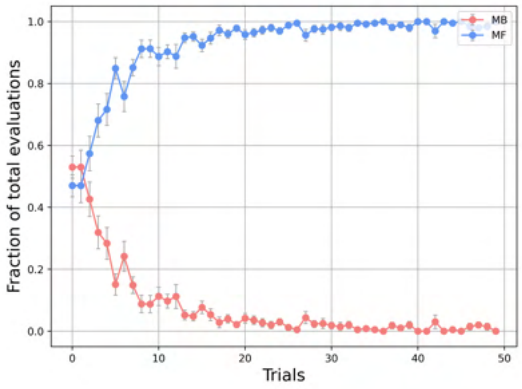

Figure 6.3 The dual process arbitration across trials for VoI based arbitration. The fraction of total evaluations are plotted for both - MB and MF controller.

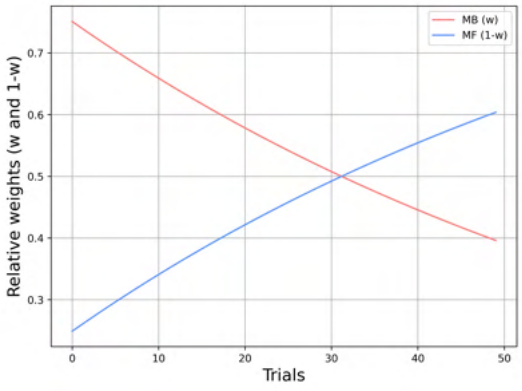

Figure 6.4 The dual process arbitration across trials for weighted-hybrid arbitration. The relative weights (w and $1 - w$) are plotted for both - MB and MF controller.

Simulation using best-fit parameters

The VoI based arbitration, weighted-hybrid arbitration, SARSA and random agents were simulated using the best-fit parameters. The simulations results are plotted.

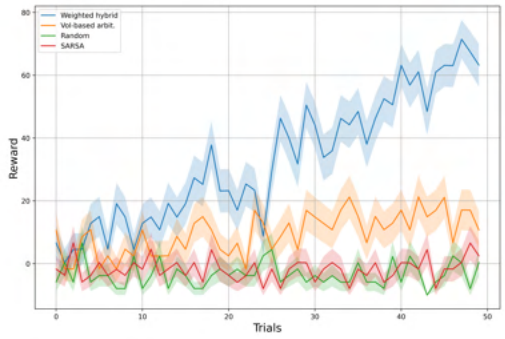

Figure 6.5 Performance comparison. The plot is generated by simulating each model with the best-fit parameters for 50 runs. The mean and SEM are plotted.

Data and Code Availability

The experimental code, analysis scripts and preprocessed data for the empirical and computational studies presented in the thesis is made available online at `https://gitlab.com/berakrishn/skill-learning-in-internally-guided-sequencing`.

तमसो मा ज्योतिर्गमय

From darkness, lead me to light.