

CHANGING MOVEMENT PATTERNS USING REINFORCEMENT LEARNING

By

Tzu-Hsiang Lin

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Kinesiology—Doctor of Philosophy

2020

## ABSTRACT

### CHANGING MOVEMENT PATTERNS USING REINFORCEMENT LEARNING

By

Tzu-Hsiang Lin

Humans interact with the world by generating movements, which make it important to understand the process of motor learning. There are two aspects of motor learning: (1) an improvement in task performance (e.g., learning to throw farther), and (2) a change in the movement pattern (e.g., learning to throw with an improved coordination or technique even if there is no change in task performance). Most studies on motor learning focus on the first aspect of task performance; however, the second aspect of movement pattern is also important and ubiquitous in our daily life - for example, we learn a better movement pattern to carry heavy objects to prevent injuries or the patients re-learn to perform movements in the rehab setting. In this dissertation, I designed a learning protocol that provided reinforcement feedback to guide participants to learn alternative movement patterns to perform the same task. Reinforcement feedback provides participants with a signal to start exploring different movement patterns but does not provide direct information about the desired movement pattern. Therefore the key question of this dissertation was to address the issue of how to schedule the reinforcement feedback to shift participants toward an alternative movement pattern in tasks requiring coordination of multiple body segments. In experiment one, I tested how providing ‘online’ reinforcement feedback (i.e. feedback provided during the movement) could shift the participants to alternative movement patterns. In experiment two, I tested how providing ‘terminal’ reinforcement feedback (i.e. feedback provided at the end of the movement) could shift participants toward alternative movement patterns, and if an adaptive method

that adjusts reinforcement based on prior performance had better learning outcomes.

In summary, I found: (1) reinforcement feedback can be used to change movement patterns in task requiring coordination of multiple body segments, although it is less successful when compared to its use in simpler tasks (2) online reinforcement feedback resulted in quick changes toward the desired movement pattern, and the amount of practice was the primary factor that determined retention, and (3) terminal reinforcement feedback resulted in less change toward the desired movement pattern, and an adaptive algorithm was needed to achieve better learning outcomes. These results contribute to the fields of motor learning and computational motor neuroscience to understand how the central nervous system uses feedback to change movement patterns, and can be applied to the fields of skill acquisition or motor rehabilitation to help people learn motor skills.

## TABLE OF CONTENTS

|                                                                                                          |     |
|----------------------------------------------------------------------------------------------------------|-----|
| LIST OF FIGURES .....                                                                                    | vii |
| CHAPTER 1 INTRODUCTION .....                                                                             | 1   |
| Focus of dissertation .....                                                                              | 3   |
| CHAPTER 2 LITERATURE REVIEW .....                                                                        | 6   |
| Learning in a redundant motor system.....                                                                | 6   |
| Data-driven approaches to describe movement patterns .....                                               | 7   |
| Task-driven approaches to describe movement patterns .....                                               | 8   |
| Learning with external feedback.....                                                                     | 9   |
| Error-based learning.....                                                                                | 10  |
| <i>The paradigm</i> .....                                                                                | 10  |
| <i>The mechanism</i> .....                                                                               | 11  |
| <i>Experimental results</i> .....                                                                        | 12  |
| <i>Optimization</i> .....                                                                                | 12  |
| <i>Summary</i> .....                                                                                     | 13  |
| Reinforcement learning.....                                                                              | 14  |
| <i>The paradigm</i> .....                                                                                | 14  |
| <i>The mechanism</i> .....                                                                               | 15  |
| <i>Exploration vs. noise</i> .....                                                                       | 15  |
| <i>Summary</i> .....                                                                                     | 16  |
| Use-dependent learning .....                                                                             | 17  |
| <i>The paradigm</i> .....                                                                                | 17  |
| <i>The mechanism</i> .....                                                                               | 17  |
| <i>Summary</i> .....                                                                                     | 18  |
| How to learn alternative movement patterns .....                                                         | 18  |
| Summary .....                                                                                            | 20  |
| CHAPTER 3 LEARNING ALTERNATIVE MOVEMENT PATTERNS USING<br>REINFORCEMENT FEEDBACK IN A REACHING TASK..... | 21  |
| Abstract .....                                                                                           | 21  |
| Introduction.....                                                                                        | 22  |
| Methods- Experiment 1.....                                                                               | 24  |
| <i>Participants</i> .....                                                                                | 24  |
| <i>Apparatus</i> .....                                                                                   | 25  |
| <i>Task</i> .....                                                                                        | 25  |
| <i>Procedure</i> .....                                                                                   | 26  |
| <i>Providing Reinforcement feedback</i> .....                                                            | 26  |
| <i>Groups and Reinforcement schedules</i> .....                                                          | 28  |
| <i>Data Analysis</i> .....                                                                               | 29  |
| <i>Statistical Analysis</i> .....                                                                        | 30  |
| Results.....                                                                                             | 31  |

|                                                                 |    |
|-----------------------------------------------------------------|----|
| <i>Trunk-hand distance - Far targets</i> .....                  | 31 |
| <i>Trunk-hand distance - Near targets</i> .....                 | 32 |
| <i>Path Length - Far targets</i> .....                          | 34 |
| <i>Path Length - Near targets</i> .....                         | 35 |
| Discussion of Experiment 1 and rationale for Experiment 2 ..... | 36 |
| Methods – Experiment 2 .....                                    | 37 |
| <i>Statistical Analysis</i> .....                               | 37 |
| Results .....                                                   | 37 |
| <i>Trunk-hand distance - Far targets</i> .....                  | 37 |
| <i>Trunk-hand distance - Near targets</i> .....                 | 39 |
| Discussion of Experiment 2 .....                                | 40 |
| General Discussion .....                                        | 40 |

#### CHAPTER 4 SHAPING REINFORCEMENT FEEDBACK TO INDUCE CHANGES IN MOVEMENT PATTERNS IN A THROWING TASK .....

|                                               |    |
|-----------------------------------------------|----|
| Abstract .....                                | 45 |
| Introduction .....                            | 46 |
| Methods: experiment 1 .....                   | 49 |
| <i>Participants</i> .....                     | 49 |
| <i>Apparatus</i> .....                        | 49 |
| <i>Task</i> .....                             | 50 |
| <i>Score feedback</i> .....                   | 50 |
| <i>Providing reinforcement feedback</i> ..... | 51 |
| <i>Procedures</i> .....                       | 52 |
| <i>Groups</i> .....                           | 52 |
| <i>Data analysis</i> .....                    | 53 |
| <i>Statistical analysis</i> .....             | 55 |
| Results: experiment 1 .....                   | 55 |
| <i>Score</i> .....                            | 56 |
| <i>Punishment rate</i> .....                  | 57 |
| <i>Trunk velocity</i> .....                   | 57 |
| <i>Hand velocity</i> .....                    | 58 |
| <i>Task space variability</i> .....           | 59 |
| <i>Null space variability</i> .....           | 60 |
| <i>Summary of experiment 1</i> .....          | 61 |
| <i>Rationale for Experiment 2</i> .....       | 61 |
| Methods- experiment 2 .....                   | 61 |
| <i>Participants</i> .....                     | 62 |
| <i>Apparatus and task</i> .....               | 62 |
| <i>Grouping and shaping methods</i> .....     | 63 |
| <i>Data analysis</i> .....                    | 64 |
| Results: experiment 2 .....                   | 65 |
| <i>Threshold</i> .....                        | 65 |
| <i>Score</i> .....                            | 65 |
| <i>Punishment rate</i> .....                  | 66 |
| <i>Trunk velocity</i> .....                   | 67 |

|                                                      |    |
|------------------------------------------------------|----|
| <i>Hand velocity</i> .....                           | 68 |
| <i>Null space variability</i> .....                  | 69 |
| <i>Summary of experiment 2</i> .....                 | 70 |
| General Discussion .....                             | 70 |
| <i>Reinforcement in multi-DOF tasks</i> .....        | 71 |
| <i>Shaping schedules</i> .....                       | 72 |
| <i>Abrupt vs. Gradual</i> .....                      | 72 |
| <i>Adaptive schedules</i> .....                      | 73 |
| CHAPTER 5 GENERAL DISCUSSION .....                   | 75 |
| Overall scope .....                                  | 75 |
| Contributions of the dissertation .....              | 76 |
| Online feedback vs delayed feedback .....            | 77 |
| Shaping reward/punishment during reinforcement ..... | 78 |
| Limitation and future direction .....                | 79 |
| Conclusion .....                                     | 80 |
| REFERENCES .....                                     | 82 |

## LIST OF FIGURES

|                                                                                |    |
|--------------------------------------------------------------------------------|----|
| Figure 3.1. Schematic of experimental setup.....                               | 26 |
| Figure 3.2. Reinforcement feedback and experimental protocol.....              | 28 |
| Figure 3.3. Variation of thresholds and actual trunk-hand distances.....       | 33 |
| Figure 3.4. Mean trunk- hand distance in far and near targets .....            | 34 |
| Figure 3.5. Schematic of path length .....                                     | 35 |
| Figure 3.6. Variation of thresholds and actual trunk-hand distances.....       | 38 |
| Figure 3.7. Group mean trunk- hand distance.....                               | 39 |
| Figure 4.1. Experimental setup .....                                           | 49 |
| Figure 4.2. Mechanism of providing reinforcement feedback.....                 | 51 |
| Figure 4.3. Design of shaping methods .....                                    | 53 |
| Figure 4.4. Definition of task space and null space .....                      | 54 |
| Figure 4.5. Change in mean score.....                                          | 56 |
| Figure 4.6. Change in mean punishment rate .....                               | 57 |
| Figure 4.7. Change in standardized trunk velocity and hand velocity.....       | 58 |
| Figure 4.8. Change in task and null space variability.....                     | 60 |
| Figure 4.9. Schematic of scoring with and without reinforcement feedback ..... | 63 |
| Figure 4.10. Design of reinforcement schedules .....                           | 64 |
| Figure 4.11. Change in trunk velocity threshold.....                           | 65 |
| Figure 4.12. Change in mean score.....                                         | 66 |
| Figure 4.13. Change in punishment rate .....                                   | 67 |
| Figure 4.14. Change in standardized trunk velocity and hand velocity.....      | 68 |

|                                                              |    |
|--------------------------------------------------------------|----|
| Figure 4.15. Change in task and null space variability ..... | 69 |
|--------------------------------------------------------------|----|

## CHAPTER 1 INTRODUCTION

Consider the situation of a beginner learning to play tennis: not only does the learner have to focus on the eventual task outcome (say hitting the ball over the net), but in order to do so, the learner has to learn to coordinate joints and limbs to hit the ball – i.e., learn a new movement pattern. This problem of learning a movement pattern is further complicated by the fact that there are many possible movement patterns to hit the ball successfully - e.g., hitting the ball with an overarm or underarm movement pattern. This is the issue of motor redundancy where the motor system has many motor solutions to perform the same task (Bernstein, 1967). However, even though multiple solutions may be available to perform the task equally well in terms of the task outcome, some solutions may be preferred to others because they have other advantages (such as efficiency or injury prevention). In these cases, an external agent (e.g., a coach) may need to shift participants from using one solution to another. The central question that this dissertation addresses is how to ‘shift’ participants from one movement pattern to another in complex motor tasks. Understanding how to best structure practice to learn such new movement patterns is an important issue not only for skill acquisition, but is also a central part of movement rehabilitation in neurological disorders like stroke.

Providing augmented feedback is a well-studied approach to guide the learner to specific movement patterns. Augmented feedback refers to feedback about the movement that is not intrinsic to the individual, and is usually provided by an external agent (coach, therapist, etc.) The learner uses this feedback as a ‘learning signal’ to modify movement patterns on future trials. Although previous approaches have distinguished based on the content of information (i.e. knowledge of results vs. knowledge of performance), a more recent distinction that has been made is based on the learning process itself – i.e., how behavior changes after providing different

types of feedback (Wolpert et al., 2011). Under this category, there are two primary types of feedback that guide the learner through different learning mechanisms: (1) error feedback (2) reinforcement feedback.

Error feedback measures the difference between the performance and the target, and thus indicates both the magnitude and the direction of the error. For example, when a tennis player undershoots the target (and can see where the ball lands), the player can use the information to hit harder toward the direction of the target on the next trial. Since the magnitude and direction of the error are known, the learner modifies movement patterns to decrease the error trial-by-trial. Error-based learning is useful when the error is non-zero, it directs the learner to one of the many movement patterns that will bring the error to zero. However, there are two limitations of error-based learning – (i) once the error is zero, there is no learning signal to modify movement patterns further, even though learning may be required to modify movement patterns in tasks with redundancy, and (ii) implementing error in terms of movement patterns involving lots of body segments can result in high-dimensional feedback that is difficult to process for the learner. For example, when a learner performing a tennis serve receives a vector of errors on the shoulder, elbow, and wrist, it is difficult to use this information to make adjustments on all three joints at the same time.

An alternative to error feedback is reinforcement feedback. Reinforcement feedback evaluates the performance and effectively provides a ‘good’ or ‘bad’ signal. In other words, it provides coarse-grained information that either signals the learner to retain the movement pattern (if the signal is good) or change the movement pattern (if the signal is bad). Unlike the systematic changes observed in error-based learning where errors can be gradually reduced to zero, reinforcement learning is characterized by exploration because the learner does not know

the size or the magnitude of the error with great precision. However, an advantage is that the reinforcement signal is low-dimensional (making it easy for the learner to use), and this process of exploration can be used to modify movement patterns even after the task error has become zero.

Exploration helps to learn alternative movement patterns but here again, the challenge with task with multiple degrees of freedom (DOF) is that the learning process could be inefficient when the learner explores along the irrelevant or incorrect dimensions. Because the reinforcement is low-dimensional, it does not directly ‘guide’ the exploration toward the desired solution. Going back to the tennis serve example, telling a participant ‘bad serve’ does not provide sufficient information on how to change the shoulder, elbow and wrist motions. One solution to this problem is through ‘shaping’ (Ferster & Skinner, 1957), – i.e. gradually manipulating feedback based on the learner’s behavior. In other words, the learner has a higher possibility to receive a ‘good’ reinforcement feedback when exploring in the right direction. However, how these shaping schedules should be used in multi-DOF tasks to shift participants from one pattern to another is not well understood. To study this question, I used a reinforcement-based algorithm with different shaping methods to guide participants to explore different solutions in an efficient way.

### **Focus of dissertation**

In this dissertation, I investigated two main questions: (1) can reinforcement feedback to change movement patterns involving multiple degrees of freedom (DOF), and (2) how can different shaping methods help to change movement patterns?

I designed a virtual throwing task based on trunk-hand coordination to study the change in movement patterns. Participants were asked to learn the coordination of trunk and hand kinematics to meet the task goal. The redundancy in the task was based on the fact that many different movement patterns could be used to achieve the task, which allowed us to examine how reinforcement could help participants shift from one movement pattern to another. In this case, participants typically did not use trunk motion at the start of performing this task, and we applied reinforcement feedback so that a ‘punishment’ (i.e., a low score) was applied when the trunk movement was below a certain ‘threshold’. I manipulated this threshold during practice using different methods to examine how participants changed their movement pattern. Participants initially performed a ‘pre-test’ with no reinforcement, then practiced with reinforcement feedback for a training period, and then performed a ‘post-test’ with no reinforcement. This allowed us to examine the change in movement pattern due to the reinforcement feedback, and the retention of this movement pattern after it was removed.

In experiment one, I examined the effect of learning alternative movement patterns with online reinforcement feedback (i.e. feedback given during the movement). Participants were grouped into an ‘abrupt’ group where the threshold changed abruptly and a ‘gradual’ group where the threshold changed gradually. Results showed that both groups learned alternative movement patterns, although the abrupt group retained this movement pattern even after the reinforcement was removed.

In experiment two, I examined the effect of learning new movement patterns with different shaping methods in a discrete task (i.e. where feedback was given only at the end of the movement). A key question to address was whether ‘adaptive’ shaping methods (i.e. where the threshold was modified according to the participants’ performance) would yield greater changes

in the movement pattern relative to open loop shaping methods (abrupt/gradual). Results showed that (i) reinforcement feedback in discrete tasks created smaller changes in the movement pattern (compared to the online reinforcement feedback), and that adaptive shaping methods which resulted in a moderate rate of reinforcement were more effective at creating changes in movement pattern relative to open loop methods.

The first contribution of this dissertation is to study reinforcement learning in the context of redundant tasks that have multiple solutions to achieve the task goal. Our results show that although reinforcement learning is quite successful when it is provided *during* the movement, reinforcement learning is considerably less successful at eliciting changes in movement patterns in discrete tasks, when it is provided at the end of the movement. The second contribution is to understand how shaping methods can be used to make reinforcement learning in discrete tasks effective. Adaptive shaping methods, which are based on the participants' performance, were found to be more effective at creating changes in movement patterns compared to open-loop methods. Overall, these results point to the need for further work in understanding reinforcement learning in real-world tasks and how they can be applied to motor learning in skill acquisition or rehabilitation.

## CHAPTER 2 LITERATURE REVIEW

The focus of the dissertation is to investigate how the change in movement patterns can be guided by using reinforcement feedback. In this context, I will use a theoretical framework from computational motor-neuroscience (Wolpert et al., 2001). This framework does not focus on *where* motor learning happens in the central nervous system but rather focuses on *what kind of computation* is implemented throughout learning (Wolpert & Flanagan, 2016). In a typical learning paradigm, the learners receive ‘learning signals’ and generate new movement patterns to perform the task (Jordan & Rumelhart, 1992). The computation happens after receiving the learning signal until the onset of the movement. Learning signals are the input to the nervous system and the movement patterns are the output (Franklin & Wolpert, 2011). In this review section, I discuss (i) the challenge of learning in a redundant system, and (ii) the different types of learning signals in motor learning in the context of learning new movement coordination patterns.

### **Learning in a redundant motor system**

Humans have the ability to perform different movement coordination patterns to perform the same task. For example, one can reach for a target with the elbow flexed or the elbow extended. Both movement patterns are equally ‘good’ from the viewpoint of achieving the task goal. This many-to-one mapping, which arises because there are more degrees of freedom (DOF) in the motor system than that are constrained by the task, was framed as the challenge of redundancy in the motor system (Bernstein, 1967). Motor redundancy can be evaluated at different levels of the motor system (e.g., motor neurons, muscle groups, kinetics, and kinematics); here, I focus on the level of kinematics. For example, to reach a target, only the

position of the hand is constrained by the task directly. Other DOF such as elbow and shoulder are not directly constrained by the task, and therefore give rise to different ways of performing the task. This flexibility is especially useful to deal with the uncertain environments, for example, we can reach for an object even when there is an obstacle in the way.

Understanding the relation between motor learning and motor redundancy is critical because that movement patterns can be variable while still achieving the same task outcome (Guigon et al., 2007; Latash, 2012; Singh et al., 2016). So, while learning results in changes in the task outcome, how do participants settle on a movement pattern or ‘shift’ to an alternative movement pattern when several movement patterns lead to the same task outcome? Before describing this process of learning in term of movement patterns, I briefly review methods that quantify movement patterns in the context of redundancy.

### **Data-driven approaches to describe movement patterns**

Human movements often have many DOFs, but the important variance usually lies in a low-dimensional space. There are many dimensionality reduction techniques to find this lower-dimensional space. One widely used technique is principal component analysis (PCA), the goal of PCA is to find the linear transformation to capture the variance of the data with fewer dimensions (represented by the first few principal components). The principal components that have high explained variance can be seen as the important dimensions that drive the movement (Gløersen et al., 2018; Witte et al., 2010). PCA is advantageous because it is easy to implement and has been successfully used to extract low-dimensional data in a variety of contexts. However, not all the data have linear patterns, especially human movements are considered nonlinear due to the anatomical constraints.

Another category of dimensionality reduction techniques is nonlinear methods that consider the nonlinearity in human movement pattern. Nonlinear methods have a better biological explanation than linear methods, it assumes the dynamics of human movement lies on a nonlinear manifold (Ficuciello et al., 2018; Jenkins & Matarić, 2004; Safonova et al., 2004; Wang & Suter, 2007). The nonlinear method groups the dimensions based on the balance between local similarity and global similarity. For example, to throw a ball, the elbow and shoulder are measured as local whereas the whole body is global. The results will show the shoulder and the elbow are grouped together to achieve the goal of reducing dimensions. There are two advantages of using nonlinear techniques: 1) It makes sense to apply nonlinear techniques to the nonlinear motor system 2) It is possible to discover the hidden dynamics in the data. However, a more powerful technique means that the lower dimensions are harder to interpret, and also the computation is complicated. Unless the linear methods cannot approximate the data well otherwise it is not advised to start from nonlinear methods.

### **Task-driven approaches to describe movement patterns**

The above mentioned dimensionality reduction approaches describe the relation between different degrees of freedom, but do not address how this relation affects task performance. Therefore, a second class of approaches measures movement patterns in terms of how they achieve the task goal. For example, consider a reaching movement: although there are multiple joints in motion (shoulder/elbow/wrist), the important point from a task goal is whether the hand gets to the target or not. The Uncontrolled manifold hypothesis (Scholz & Schöner, 1999) adapted this concept and divided the task into two independent spaces: a “task space”, where the dimensions that affect task performance, and a “null space”, which is the dimension along which

there is no change in the task performance (Latash et al., 2002; Scholz et al., 2002). For example, in the reaching context, the ‘task space’ would lead to changes in hand position, whereas the ‘null space’ would lead to no changes in hand position. This method has also been adopted for other methods (Cusumano & Cesari, 2006; Müller & Sternad, 2004).

From a motor learning standpoint, the task and null spaces play very different roles. Because variations along the task space causes change in the outcome, learning should result in a reduction of the task space variability to have stable performance. However, there is no such constraint on the null space as, by definition, it does not influence task performance in any way. On the one hand, null space variability may decrease with learning as a way of being more consistent and finding an ‘optimal’ solution. However, the null space variability could also increase with learning as it provides flexibility in performing the task with different movement patterns. For example, if some of the solutions are riskier to cause injuries due to extreme body postures, the null space provides a way to ‘shift’ coordination pattern and move to a better one without affecting task performance. Addressing this kind of motor learning, which does not focus on the change of performance but rather on the change of movement patterns, is the main focus of the dissertation. Because this process is complex and usually takes a long time (e.g. pro tennis players change their stroke technique), understanding the computational aspects of motor learning can help to design a paradigm to change movement patterns (Neilson, 1993; Schaal & Schweighofer, 2005).

### **Learning with external feedback**

This review section is based on the concept that the nervous system receives external feedback of the task as the input, and processes this information to update the movement

coordination pattern (Cusumano & Cesari, 2006; Müller & Sternad, 2004). The purpose is to understand what types of input signals guide motor learning and how humans learn with different types of feedback. There are three forms of learning paradigms that provide different types of feedback and lead to different learning mechanisms (Wolpert et al., 2011): (1) error-based learning; (2) reinforcement learning; (3) use-dependent learning.

First, I introduce the paradigm of each learning form and describe the details about the mechanism. Second, I discuss the question of learning alternative movement patterns in each mechanism. Third, I discuss future directions in the motor learning research. I separate the definition of the learning paradigm and learning mechanism due to the psychological aspects of human learning (Jordan & Rumelhart, 1992). The learning paradigm is defined by the structure of the feedback, and the learning mechanism is related to how the learner responds to the feedback.

## **Error-based learning**

### ***The paradigm***

In the error-based learning paradigm, the environment provides a signed feedback signal to the learner. The learner corrects the movement based on the magnitude and the sign of the feedback. The magnitude shows how much to correct and the sign shows which direction to correct. For example, in a dart-throwing task, the position of the dart shows the direction and the distance to the bullseye, this error information guides movement coordination pattern in the next trial. Error-based learning paradigm was studied by adaptation task, e.g. reaching in force fields (Bhushan & Shadmehr, 1999), visuomotor rotation (Krakauer et al., 2000), and prism goggle

adaptation (Martin et al., 1996). The experimenters introduced perturbations to create errors to show how participants correct the errors.

### ***The mechanism***

Providing feedback with the direction and magnitude in the task space triggers an error-based learning mechanism. After receiving the feedback, the learner compares the predicted outcome to the feedback to calculate the error. After certain trials of practice, the learner will be able to associate the movement patterns to the gradient of the error, i.e. how the error changes after modifying the movement patterns. The learner then modifies movement patterns based on the direction of the gradient to minimize the error trial-by-trial until the error approaches to zero (Wolpert et al., 2011). The predicted outcome is generated by the forward internal model (Jordan & Rumelhart, 1992; Wolpert et al., 1995). A good analogy of this process is fitting the parameters of the model with a least-squares algorithm. Each new trial is a new data point in the system, the learner runs the algorithm to update the parameter for the next prediction. Error-based learning is a type of model-based learning (Haith & Krakauer, 2013). This mechanism is effective but computationally heavy due to the process of finding the gradient and minimizing the error. Error-based learning is not easy to implement when learning a complex movement – for e.g., imagine receiving errors of hip, knee, ankle, and toe when practicing a gymnastic move, the learner may not be able to effectively use this information to update the next movement. Therefore, addressing how error-based learning can be used in complex tasks remains an important issue in computational motor learning.

## ***Experimental results***

Error-based learning mechanisms are typically studied using motor adaptation tasks. In this type of task, participants experienced perturbations while performing a well-developed motor skill. The goal is to see how participants modify movement coordination to overcome the perturbation. For example, in a visuomotor rotation task (Tseng et al., 2007), participants controlled one cursor from home to the target position under a visual rotation i.e. the visual feedback deviates from actual hand trajectory. To successfully reach under +45 degrees of rotation, the participants needed to aim (the predicted outcome from the forward model) for -45 degrees. The error was calculated between the aiming angle and the visual feedback, participants adapted the aiming angle from 0 to -45 degrees trial-by-trial. By using a state-space model to describe trial-by-trial changing of aiming angle, the results showed the adaptation was driven by the prediction error, not the target error. Similar results were shown in other studies (Criscimagna-Hemminger et al., 2010), the error-based mechanism is robust across non-redundant tasks.

## ***Optimization***

Another question that is not mentioned in error-based learning is how humans modify the movement coordination pattern based on the gradient of the error. Studies showed motor learning can be seen as an optimization process (Körding & Wolpert, 2004; Selinger et al., 2015), but whether humans follow specific algorithms is still unclear. One possible optimization algorithm in human learning is gradient descent. This is a heuristic algorithm to find the movement pattern that changes the performance the most. For example, to learn a tennis forehand stroke, the learner tries different ways to hit the ball for many trials. Then the learner found modifying

elbow angle brings the greatest change of the speed of the ball. With this information the learner focuses on modifying the elbow angle to get the best performance. This concept was shown in a lab experiment: participants modified the effector with artificial noise in a redundant task (Wolpert et al., 2011). The artificial noise created a huge change to the performance, therefore the participants tried to minimize the errors from this effector.

Gradient descent provides a good framework to describe how human associate the error to the movement pattern. However, searching through all the possible movement patterns is not feasible when there are too many possible solutions. A modified version of gradient descent seems more probable: stochastic gradient descent (SGD). Instead of searching through the entire variable space, SGD only samples a portion of the variables. The learning curve of SGD looks similar to the human motor learning curve. Both learning curves show different variability structures in different stages of learning. Motor learning and optimization algorithms share similar research interests in memory, variability, learning rate, and learning steps. It is possible to use ideas from SGD research to show how humans optimize errors during motor learning (Körding & Wolpert, 2004).

### ***Summary***

Error-based learning is a simple and effective mechanism for learning new motor skills. It has the advantage of faster learning speed and the disadvantage of harder to reach alternative coordination. It has been shown that humans use this mechanism in many motor adaptation tasks. However, the effectiveness of error-based learning in a higher-dimensional tasks are still not very clear, providing errors is problematic when there are too many dimensions that need to be corrected.

## **Reinforcement learning**

### ***The paradigm***

The concept of reinforcement learning paradigm stems from the research of operant conditioning in behavioral psychology. The theory focuses on the responses after the organism receives stimulus from the environment, with two types of stimuli associated with the responses- reinforcement and punishment (Skinner, 1938). Reinforcement can be any stimulus that strengthens the responses and punishment is any stimulus that weakens the responses. The organism learns through either maximizing future reinforcement or minimizing future punishment. Later on, researchers proposed different schedules and different types of reinforcement feedback to consolidate behavior (Reynolds, 1961). To make the concept approachable, reward and punishment learning is widely used to explain operant conditioning.

With the development of machine learning and artificial intelligence, computer scientists designed the algorithm of reinforcement learning from the concept of interaction between the organism and the environment (Sutton & Barto, 2017). The concept of reinforcement learning is to map the state of the environment to the actions by maximizing predicted rewards. To construct a reinforcement learning paradigm, the environment provides unsigned feedback to the learner, often an overall evaluation of the performance. The feedback can be as simple as success or failure, or as complex as numbers from abstract math functions (Wolpert et al., 2001). Studies also provided graded feedback like monetary rewards to study how the brain learns through reward prediction (Galea et al., 2015). The reinforcement learning paradigm is a natural way to learn in the real world. Receiving the feedback of success or failure reinforces the behavior by exploring possible actions.

### ***The mechanism***

In contrast to error-based learning, the learner does not receive signed feedback of the task. Therefore, the learner has no direction of modifying movement coordination from binary feedback. A successful reinforcement learning mechanism includes two aspects - exploration and exploitation. Exploration searches the solution space and exploitation reproduces the coordination once the system has found a good solution. The long-term goal is to maximize the possibility of getting good feedback. One challenge is the ratio between exploitation and exploration. Exploitation ensures that the learner maintains current performance but doesn't help to improve the performance. Exploration provides the chance to improve the performance but the learner may not explore in the correct dimension, resulting a worse performance. As a result, reinforcement learning is often slow because of the uncertainty in the exploration, especially when the learner fails to explore the correct motor solutions. Since the forward model is not directly involved in this process, the reinforcement learning mechanism is a type of model-free learning (Haith & Krakauer, 2013).

### ***Exploration vs. noise***

Traditionally movement variability was seen as ‘noise’, and therefore considered detrimental to motor learning. However, because variability can also be due to exploration, there is evidence to show that variability has a positive role in motor learning (Dhawale et al., 2017; Herzfeld & Shadmehr, 2014; Murillo et al., 2017; Newell & Corcos, 1993; Wu et al., 2014). The variability is task-related, because of exploration, not the noise in the motor system. When the learner actively explores different movement patterns, the variability will be high in both task

space and joint space. However, these two sources of variability are hard to identify in the experiment.

One solution to separate exploration and noise is to assume that participants are rational in the experiment, and only explore when they receive unsatisfactory outcomes. So the variability after ‘good’ trials can be defined as noise, and the learning process can be modeled with a Kalman filter (Therrien et al., 2016). However, this assumption of a ‘rational’ learner becomes more difficult when participants are engaged in high-dimensional tasks, where the distinction between exploration and noise becomes less obvious. Moreover, traditional linearized metrics like variance become more difficult to interpret in higher dimensions and also ignore the temporal component of the variation (Stergiou & Decker, 2011). Therefore, it is critical to study high-dimensional tasks to fully study the relationship between variability and exploration.

### ***Summary***

Reinforcement learning is slower to reach the solution space but it has the potential to learn alternative movement coordination patterns based on the amount of exploration. Reinforcement learning provides a flexible paradigm where experimenters can reduce the information in high dimensional joint space to binary feedback, and as a result it provides a tool to study learning of complex movements.

## **Use-dependent learning**

### ***The paradigm***

Use-dependent learning is a type of unsupervised learning where there is no external feedback provided to the learner. The learner relies on the internal feedback and experience to learn the way to solve the task.

### ***The mechanism***

Use-dependent learning is the phenomenon that humans learn by repeating the same movement even though there is no external feedback provided (Krakauer & Mazzoni, 2011; Wolpert et al., 2011). This was shown in a reaching task that participants reduced variability by practicing the same task many times and caused a biased angle shift when reached to neighbor targets (Schaal et al., 2003). Moreover, since repeating the same task is what learners do to learn a new motor skill, use-dependent learning mechanism can happen along with error-based learning (Diedrichsen et al., 2010). Use-dependent learning does not actively help to learn alternative movement coordination patterns. Instead, this hinders people from moving to another movement coordination pattern after movement coordination pattern is well developed. This mechanism can be seen as a form of unsupervised learning (Jordan & Rumelhart, 1992; Todorov & Ghahramani, 2003). When the learner does not receive an external learning signal, all the available information for learning is based on the sensory input. The learner builds a probability model from all the sensory input then selects the most probable one as the next coordination. When one specific coordination is reproduced more than the others, it has a higher possibility. Then the learner has more chances to reproduce that coordination again. This explains why participants repeat the movement coordination pattern in the previous trials.

## ***Summary***

Use-dependent learning is not an independent mechanism from error-based or reinforcement-based learning. Studies showed it happens along error-based or reinforcement-based learning (Todorov & Ghahramani, 2003). For example, when the learner experiences a series of trials with zero error, from the error-based mechanism we should not observe much coordination change because zero error means there is nothing to be improved. Then use-dependent mechanism takes over the learning process where the learner tends to stick to the coordination in the previous trial. Another example in reinforcement learning, after a series of good feedback, the learner tends to keep to the movement pattern that has the largest possibility to receive good feedback. Use-dependent learning is an important mechanism to consider but often ignored when discussing the learning process.

## **How to learn alternative movement patterns**

This dissertation focuses on how to use a different movement pattern to solve the same task. The aforementioned three learning paradigms are all possible to guide the learners to change the movement patterns. However, one might be more feasible for certain tasks. Understanding the difference between all three paradigms from the perspective of computation helps to design proper learning protocol.

Error-based learning can help the learner to approach to the solution space, but hard to move to alternative coordination (Wolpert et al., 2011). This is because of all of the possible movement patterns in the task space have similarly low error. The error gradient directs learners to the solution space, but not within the solution space, so the mechanism of moving to alternative movement patterns is weak. An alternative is to provide the learning signal directly in

the joint space instead of the task space. This type of learning in skill acquisition research called observational learning, the learner observes a “good” movement pattern and try to imitate the pattern. Similarly, in robotics research, imitation learning is widely applied to teach a motor function to the robot (Schaal et al., 2003). Imitation learning provides the learning signal on coordination level, therefore the learner can compare prediction outcome to the imitation signal to construct an error-based learning paradigm. Imitation learning is faster than other learning algorithms in robotics. If motor learning researchers borrow this concept to design the learning signal in joint space, how the brain handles error feedback in a higher-dimensional space could be studied.

Reinforcement learning, on the other hand, places a big emphasis on exploration. The learner explores actively to maximize future reward (or minimize future punishment). But it might be inefficient when the learner explores the whole solution space. One way to improve efficiency is to implicitly inform the nervous system which end-effector is important. For example, researchers added artificial noise to one of the end-effectors. Consequently, participants modified coordination by minimizing the noise in the noisy end-effector (Mehler et al., 2017; Thorp et al., 2017). In another study, the experimenters designed closed-loop reinforcement feedback where the current feedback was defined by previous feedback (Therrien et al., 2016). This design makes sure participants explore the same direction to ensure the learning performance. It is possible to reinforce a specific coordination pattern when the feedback was provided on the level of joint space.

The idea of using error-based learning to move to the solution space then use reinforcement learning to change the movement pattern was proposed as a hypothesis (Wolpert et al., 2011) – however, there is no empirical evidence to show the combination will be effective.

But compared to the three learning paradigms, reinforcement learning is more likely to guide the learner to change movement patterns due to the active exploration. The learner sees the whole movement pattern as one action, the goal is to find the action which brings the highest future rewards. Although there is no empirical evidence to support the brain is using this algorithm to learn new movement patterns, in machine learning reinforcement learning, had great success in learning complex tasks. Therefore, the main focus of my work will use reinforcement learning.

## **Summary**

Studying learning new movement patterns is a complex procedure with two main steps. First, there is a need to select a method to reduce the dimensions of the motor system, and make it easier to understand the change in patterns during learning. These dimensionality methods can be linear or nonlinear. Second, there is a need to understand how different learning mechanisms guide the learner toward new movement patterns. Based on these learning mechanisms, the focus of this dissertation is to use reinforcement learning to alter movement patterns.

## **CHAPTER 3 LEARNING ALTERNATIVE MOVEMENT PATTERNS USING REINFORCEMENT FEEDBACK IN A REACHING TASK**

### **Abstract**

One of the characteristic features of the human motor system is redundancy– i.e., the ability to achieve a given task outcome using multiple movement patterns. However, once participants settle on using a specific movement pattern, the process of learning to use a new alternative movement pattern to perform the same task is still poorly understood. Here, using two experiments, we examined this process of how participants shift from one movement pattern to another using different reinforcement schedules. Participants performed a virtual reaching task, where they moved a cursor to different targets positioned on the screen. Our goal was to make participants use a movement pattern with greater trunk motion, and to this end, we provided reinforcement by making the cursor disappear if the trunk motion during the reach did not cross a specified threshold value. In Experiment 1, we compared two reinforcement schedules in two groups of participants– an abrupt group, where the threshold was introduced immediately at the beginning of practice, and a gradual group, where the threshold was introduced gradually with practice. Results showed that both abrupt and gradual groups were effective in shifting their movement patterns to involve greater trunk motion, but the abrupt group showed greater retention when the reinforcement was removed. In Experiment 2, we examined the basis of this advantage in the abrupt group by using two additional control groups. Results showed that the advantage of the abrupt group was because of a greater number of practice trials with the desired movement patterns. Overall, these results show that reinforcement can be successfully used to shift movement patterns, which has potential in the rehabilitation of movement disorders.

## Introduction

Motor redundancy – the ability to perform a given motor task using different movement patterns - is a hallmark of the human motor system. Given the large number of degrees of freedom (DOF) in the body, the question of how the nervous system learns to harness this redundancy to produce goal-directed movement continues to be a central question in motor learning (Bernstein, 1967). This feature of motor redundancy (also see ‘motor abundance’, Latash, 2012) not only provides flexibility in performing everyday movements, but also underlies the phenomenon of ‘compensatory movements’ that are often observed after neurological injury such as stroke (Cirstea & Levin, 2000; Levin et al., 2009), where an alternative movement pattern is used to compensate for a specific movement deficit -e.g., using trunk motion during reaching to compensate for inadequate elbow extension. Therefore, from both theoretical and applied perspectives, it is critical to understand how participants learn to use motor redundancy.

However, in spite of the extensive focus on redundancy at several levels of the motor system in the literature (Dingwell et al., 2010; Latash, 2012; Müller & Sternad, 2004; Todorov & Jordan, 2002), there has actually been surprisingly little attention in understanding how participants learn a new alternative movement pattern to perform the same task by reorganizing the DOFs. On the one hand, there is evidence that there is significant reorganization of DOFs during motor learning of novel tasks (Konczak et al., 2009; Newell, 1986; Vereijken et al., 1992) - however these studies typically do not address the issue of whether the new movement patterns that emerge with learning are ‘alternatives’ to the initial movement pattern (i.e. whether they could be used to produce the same task outcome). On the other hand, several studies have shown that participants can quickly change to alternative movement patterns to maintain the same task

outcome (Diedrichsen, 2007; Martin et al., 2011) – however these studies have typically employed already well-learned tasks such as reaching and force production. Therefore, in order to bridge this gap, we need to understand the acquisition of a new movement pattern in a task which is an ‘alternative’ to a pre-existing movement pattern (i.e. both can produce the same task outcome), and relatedly, how these changes in movement patterns can be elicited using augmented feedback.

Here, we explored the use of reinforcement feedback as a tool, and tested different reinforcement schedules to alter movement patterns with multiple degrees of freedom. Reinforcement feedback, often summarized as ‘reward and punishment’ (Sutton & Barto, 2017; Wolpert et al., 2001), provides a scalar feedback about the movement, without providing precise error information. The goal of the participants is therefore to simply act in a way that maximizes the reward (or minimizes the punishment). It is important to note that the term ‘reinforcement’ used in this context is related but somewhat distinct from that used in the classic psychology literature (Ferster & Skinner, 1957) in that it can refer to both reward or punishment-like feedback. The rationale behind using reinforcement learning is that it is particularly well suited to the learning of novel movement patterns both from theoretical and practical viewpoints. From a theoretical viewpoint, it is different from error-dependent learning (Wolpert et al., 2011) in that there is no indication of the error magnitude, and therefore requires exploration to find the optimal solution. From a practical viewpoint, reinforcement feedback is often much simpler to provide in multi-DOF tasks because it is a simple scalar whereas in contrast, error-dependent feedback would have to be multidimensional since feedback would have to be provided on both the magnitude and direction of the error. However, it is unclear how reinforcement feedback is best utilized to elicit changes to coordination, so we examined how the reinforcement schedule

(abrupt vs. gradual) influenced the learning of a new movement pattern. Although several studies on motor adaptation (based on error-dependent learning), show greater retention (as measured by the ‘after-effects’ of adaptation) in the gradual reinforcement schedule (Kagerer et al., 1997; Shadmehr et al., 2010), we aimed to examine this relation in a reinforcement learning of a novel movement pattern.

In this study, we examined the trunk-arm coordination during reaching. This is a system with redundancy because the position of the hand in space is influenced both by configuration of the arm, as well as the configuration of the trunk. Although several previous studies have examined how participants exploit the redundancy in multiple degrees of freedom during reaching, we focused on making participants shift from a ‘typical’ reaching motion to learning a ‘compensatory’ reaching motion. Typical reaching motions in unimpaired individuals involve little to no trunk motion for targets within arm length, whereas ‘compensatory’ movement patterns (similar to that seen after stroke) involved reaching with greater trunk motion (Cirstea & Levin, 2000). We hypothesized that (i) reinforcement feedback can shift movement pattern during reaching, and (ii) participants would show greater retention of the compensatory movement pattern in the gradual group compared to the abrupt group.

## **Methods- Experiment 1**

### ***Participants***

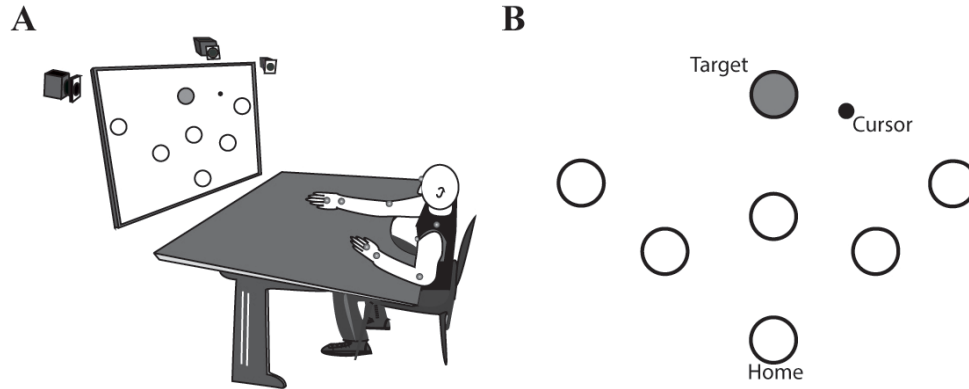
Twenty-four college students (Mean age  $\pm$  SD:  $21 \pm 1$  years, 16 Female, 4 left-handed) with no history of neurological or musculoskeletal injury participated in the experiment for extra course credit. Participants provided informed consent and all procedures were approved by Institutional Review Board at Michigan State University.

### ***Apparatus***

Participants sat in front of a desk facing a 50" (127 cm) television screen (Figure 3.1A). A motion capture system was used to record kinematics at a sampling rate of 120 Hz (Motion Analysis Corporation, Santa Rosa, CA). Ten retro-reflective markers were attached to the body - forehead, sternum, and bilaterally at the shoulder, elbow, wrist, and hand (third metacarpophalangeal joint). An additional eleventh marker was placed on the left side of the chest to distinguish left and right sides of the body.

### ***Task***

The task was a virtual reaching task where participants moved their right hand and attempted to move a cursor on a screen to specific targets. A MATLAB program was designed to receive (x,y,z) coordinates from motion capture system, and we mapped the (x,y) coordinates (corresponding to horizontal plane) of the right hand to a cursor on the screen. Six virtual targets were placed in three different directions (0, +45°, -45°) and two distances (near targets: 160 mm and far targets: 320 mm) from the starting position (Figure 3.1B). Targets were fixed at the same distances for all participants (i.e. they were not scaled to participant height or arm length). Participants were instructed to move the cursor from home position to the target to perform a virtual reaching task as fast as possible. Each trial ended when the cursor entered the specified target (and stayed inside the target for 500 ms). To keep the participants motivated throughout the experiment we provided additional feedback - if they reached the target within a target movement time (movement time had to be below 800 ms), the target was highlighted in yellow and they heard a rewarding tone, otherwise the target was highlighted in red.



**Figure 3.1. Schematic of experimental setup** (A) Participants sat in front of TV screen and their kinematics were recorded using a motion capture system. (B) Participants performed “virtual reaching” movements where they had to move the cursor on the screen to one of six different targets shown (three near and three far). The cursor position was controlled by the x-y coordinates of the marker on the right hand.

### ***Procedure***

There were a total of 7 blocks in the experiment - a pre-test, 5 blocks of training, and a post-test (Figure 3.2C). Participants performed 60 trials in each block (10 trials to each target), for a total of 420 trials in the entire experiment. In the pre-test and post-test blocks, there was no reinforcement feedback, which meant that participants could perform the reaching task with any movement pattern. In the training blocks, we introduced reinforcement feedback to constrain the participants to use a compensatory movement pattern (see Groups and reinforcement schedules section below for how this was implemented). Each trial consisted of a single outward reach from the home target to one of the six peripheral targets. The order of target presentation was randomized with the constraint that all six targets had to be presented before a target could repeat. Each experiment lasted for about 40 minutes.

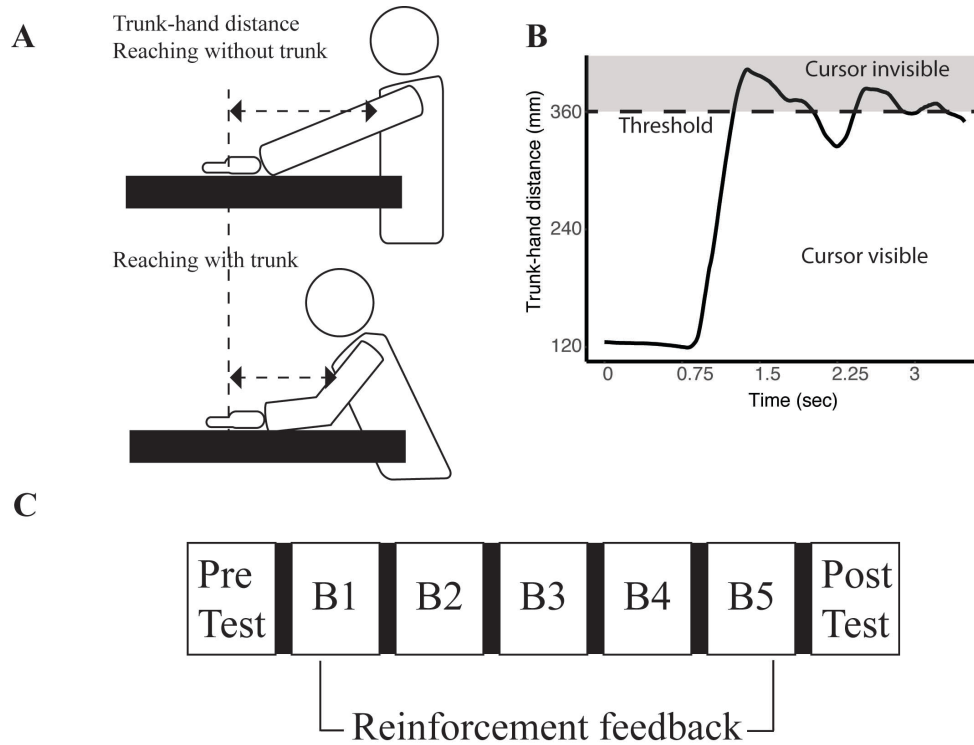
### ***Providing Reinforcement feedback***

To make participants learn an alternative movement pattern – i.e., to use more trunk motion during the reach, we provided binary reinforcement feedback by manipulating the vision

of the cursor. First, we defined the movement to be compensatory by comparing the trunk-hand distance (i.e. the distance between the sternum and hand marker measured along the horizontal plane) to a specified threshold value (Figure 3.2A). Using this threshold, we manipulated the visual feedback such that the cursor was visible as long as the trunk-hand distance was smaller than threshold (i.e. greater trunk motion), but the cursor disappeared when the trunk-hand distance was larger than the threshold (Figure 3.2B). Depending on the trunk-hand distance, the cursor could appear and disappear multiple times during a single trial, and a trial could be completed only if the cursor was visible when it was inside the target.

The instructions given to the participants were as follows: During the pre-test, participants were only given the instructions related to the task itself (i.e. moving the cursor to the targets as fast as possible). During the reinforcement blocks, participants were told that the cursor might occasionally disappear, and when this happened, they were simply instructed to explore different ways of moving their body until the cursor reappeared on screen. It is important to note that the instruction to ‘explore’ was meant only to prevent participants from stopping when the cursor disappeared - no explicit instruction was given about moving the trunk. Prior to the start of the post-test, participants were not explicitly told that the reinforcement feedback had been removed.

The disappearance of the cursor served as reinforcement feedback because participants effectively knew when they made an “error” (the error in this context was defined as not using sufficient trunk motion during the reach), but did not have any precise information about how much this error was, or how much more trunk motion was needed to make in order to get the cursor back. Participants could also not ignore this feedback because they could not complete the trial without visual feedback of the cursor.



**Figure 3.2. Reinforcement feedback and experimental protocol** (A) Reinforcement feedback was provided based on the trunk-hand distance to elicit a movement pattern with greater trunk motion. (B) The cursor on the screen was visible as long as the trunk-hand distance was smaller than a specified threshold, but became invisible when the trunk-hand distance exceeded the threshold. (C) Participants performed 60 trials for each block and total of 7 blocks. Pre-test and post-test block had no reinforcement feedback. Blocks 1-5 (B1-B5) had reinforcement feedback.

### ***Groups and Reinforcement schedules***

We tested two types of reinforcement feedback schedules - abrupt, and gradual by assigning participants into one of two groups ( $n = 12/\text{group}$ ). In the abrupt group, the threshold was immediately reduced from 600 to 360 mm in block 1, and raised back up to 600 mm in the post-test (Figure 3.3A). In contrast, in the gradual group, the threshold was gradually reduced over a set of 60 trials (a decrease of 48mm every 12 trials) so that the threshold was 360 mm only in block 2 (Figure 3.3B). Also, the threshold was raised back up gradually starting in block 5 so that the threshold was back to 600 mm by the end of block 5.

The rationale for the specific values of 600 mm and 360mm were as follows - the 600 mm was essentially large enough to be a “no threshold” condition - i.e., the trunk-hand distance

did not cross this threshold during typical reaching movements, and participants could reach all targets using their typical movement pattern. However, when the threshold was set to 360 mm, the feedback depended on whether participants reached for the far or near targets. For the three far targets, making typical reaching movements crossed this threshold (i.e. the cursor would disappear), and therefore participants were required to reorganize their movements to these targets. However for the three near targets, the threshold was still large enough that participants could use the typical reaching movements, and therefore no reorganization was required for the near targets. It is important to note that because we used fixed target distances, the thresholds mentioned above were also fixed (i.e., not body-scaled) but they were set so that most participants would have to reorganize their movement pattern when the threshold was set to 360mm.

### ***Data Analysis***

#### Trunk-hand distance

To examine changes in trunk-arm coordination, we computed the distance between sternum and right hand at the instant when the participants reached the target. This was the same variable that was used to control the reinforcement feedback. Since the hand had to travel the same distance, a longer trunk-hand distance meant lesser trunk movement, and a shorter distance meant more trunk movement.

#### Path length

To examine the temporal coordination between the trunk and the hand (i.e. whether the trunk and the hand moved simultaneously or sequentially), we used the path length. First we plotted the hand displacement (projected in the direction of the target) against the sternum displacement (also projected in the direction of the target). The path length was normalized by

dividing the actual path length by the shortest length between the start and end of the reach. Higher values indicated greater exploration, whereas lower values (closer to 1) represent a coordinated strategy with simultaneous motion of the trunk and hand.

### ***Statistical Analysis***

#### Trunk-hand Distance

Because our threshold required compensatory movements to the far targets, but not the near targets, we separated the analysis of the far and near targets. For each of these, we were specifically interested in three different comparisons: First, to examine if reinforcement feedback had an effect on the coordination (i.e. the manipulation check), we compared the pre-test block and the first ‘full threshold’ block where the threshold was set at 360 mm (i.e. block 1 for the Abrupt group, and block 2 for the Gradual group). Second, to compare how participants adapted to the reinforcement feedback, we compared the first full threshold block (block 1 for Abrupt, block 2 for Gradual) to the last full threshold block (block 5 for Abrupt, block 4 for Gradual). Third, to examine retention of the pattern after the threshold was removed, we compared the pre-test block to the post-test block. The analyses were subject to a block (2) x group (2) repeated measures ANOVA with block as the within-subjects factor, and group as the between-subjects factor. The significance level was set at .05.

#### Path length

Because the path length was a measure of exploration (i.e. how participants reorganized the motion of their DOFs), we were primarily interested only in the phase when the reinforcement feedback was on. We compared the path length between the first full threshold block and the last threshold block to show how participants developed strategies to perform the

task. We ran a block (2) x group (2) repeated measures ANOVA with block as the within-subjects factor, and group as the between-subjects factor. The significance level was set as .05.

## Results

### *Trunk-hand distance - Far targets*

The trunk-hand distance for the far targets is shown for the abrupt (Figure 3.3C) and the gradual group (Figure 3.3D). The average change between the groups is shown in Figure 3.4A. Effect of reinforcement feedback. There was a significant decrease in the trunk-hand distance from the pre-test to the first full threshold block (main effect of block:  $F(1,22) = 278.09$ ,  $p < .001$ ). In addition, the abrupt group showed a lower trunk-hand distance than the gradual group (main effect of group:  $F(1,22) = 5.17$ ,  $p = .033$ ). There was no significant Block x Group interaction ( $F(1,22) = 0.92$ ,  $p = .347$ ).

Adaptation to reinforcement. There was a significant decrease in the trunk-hand distance from the first full threshold block to the last threshold block (main effect of block:  $F(1,22) = 9.62$ ,  $p = .005$ ). In addition, the abrupt group showed a lower trunk-hand distance (main effect of group:  $F(1,22) = 7.65$ ,  $p = .011$ ). There was no significant Block x Group interaction ( $F(1,22) = 4.17$ ,  $p = .053$ ).

Post-test Retention. There was a significant decrease in the trunk-hand distance from the pre-test block to the post-test block (main effect of block:  $F(1,22) = 44.40$ ,  $p < .001$ ). In addition, the abrupt group showed a lower trunk-hand distance (main effect of group:  $F(1,22) = 8.62$ ,  $p = .007$ ), which was mediated by a significant Block x Group interaction ( $F(1,22) = 6.59$ ,  $p = .018$ ). Post hoc analysis showed there was no difference for both groups in pre-test ( $p = .467$ ), but abrupt group had smaller trunk-hand distance in post-test ( $p = .023$ ).

### ***Trunk-hand distance - Near targets***

The trunk-hand distance for the near targets is shown for the abrupt (Figure 3.3E) and the gradual group (Figure 3.3F). The average change between the groups is shown in Figure 3.4B.

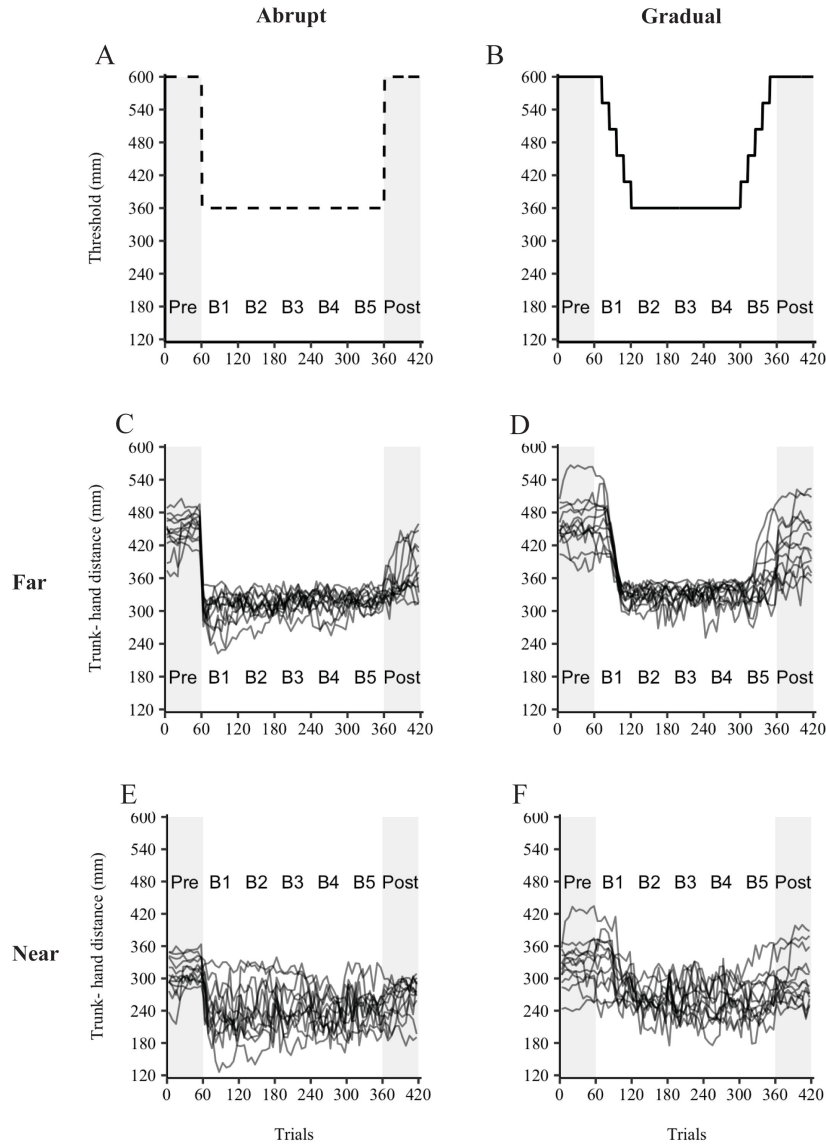
For the near targets, there was no ‘requirement’ to change movement coordination since the targets were close enough that the trunk-hand distance would be under the 360 mm threshold.

This is also seen in Figure 3.4C, which shows the proportion of time that the cursor was visible during the trial – there was a drop in the proportion for far targets initially in block 1 (indicating that the trunk-hand distance had exceeded the threshold), whereas the near targets were almost always visible.

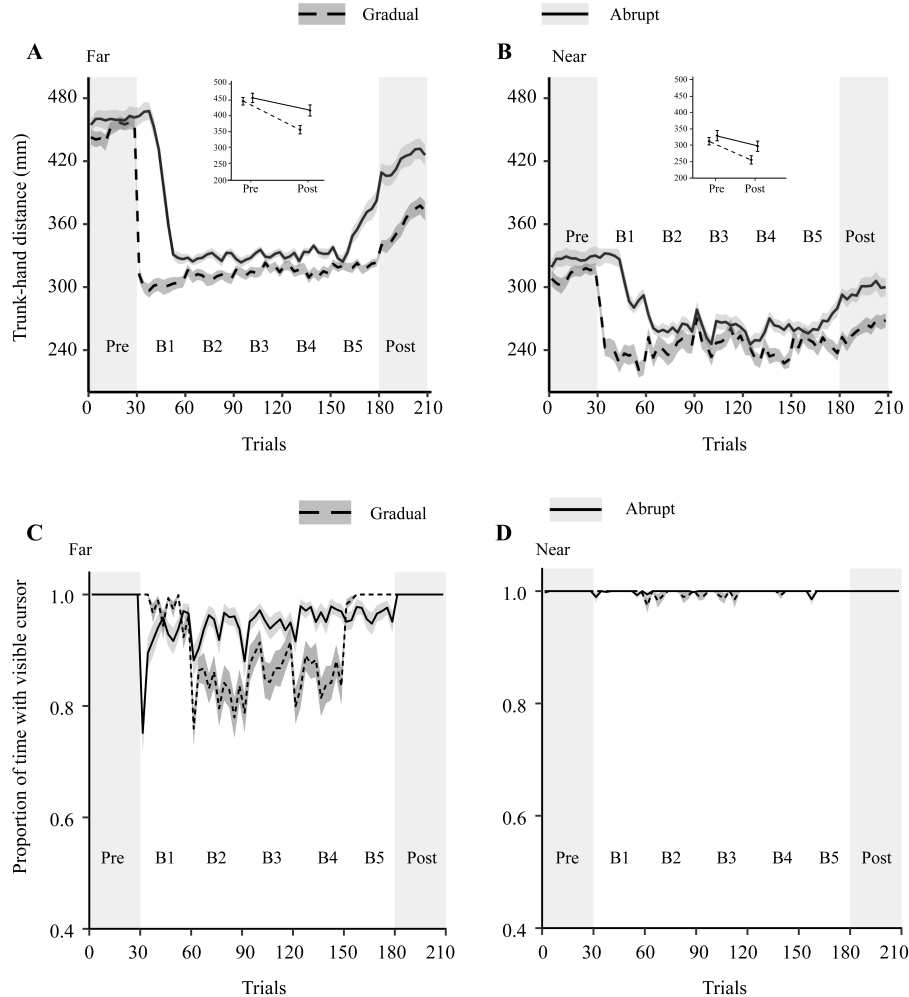
Effect of reinforcement feedback. There was a significant decrease in the trunk-hand distance from the pre-test to the first full threshold block (main effect of block:  $F(1,22) = 51.04$ ,  $p < .001$ ). There was no main effect of group ( $F(1,22) = 4.02$ ,  $p = .058$ ), and no Block x Group interaction ( $F(1,22) = 0.15$ ,  $p = .702$ ).

Adaptation to reinforcement. There was no main effect of block ( $F(1,22) = 0.12$ ,  $p = .735$ ), and no main effect of group ( $F(1,22) = 2.54$ ,  $p = .125$ ), and no Group x Block interaction ( $F(1,22) = 0.82$ ,  $p = .374$ ).

Post-test Retention. There was a significant decrease in the trunk-hand distance from the pre-test block to post-test block (main effect of block:  $F(1,22) = 18.16$ ,  $p < .001$ ). In addition, the abrupt group showed a smaller trunk-hand distance (main effect of group:  $F(1,22) = 5.73$ ,  $p = .026$ ). There was no Block x Group interaction ( $F(1,22) = 1.38$ ,  $p = .252$ ).



**Figure 3.3. Variation of thresholds and actual trunk-hand distances** (A) Threshold of abrupt group. (B) Threshold of gradual group. (C) Abrupt group at far targets. (D) Gradual group at far targets. (E) Abrupt group at near targets. (F) Gradual group at near targets. In panels (C-F), each line represents a single participant in that group. There was a marked decrease in the trunk-hand distance when the reinforcement feedback was provided (B1-B5).



**Figure 3.4. Mean trunk- hand distance in far and near targets** (A) Mean trunk- hand distance of all participants in abrupt and gradual groups when reaching to far targets. (B) Mean trunk- hand distance of all participants in gradual group when reaching to near targets. (C)(D) Mean proportion of time that cursor was visible when reaching to far and near targets. Error bars represent 1 SEM (between-participant). The gradual group showed poorer retention of the new movement pattern compared to the abrupt group in the post-test.

### ***Path Length - Far targets***

A sample trial from a single participant in each of four blocks of learning is shown in Figures 3.5A-D. The path length for the far targets is shown in Figure 3.5E.

Adaptation to reinforcement. There was a significant decrease in path length from the first full threshold block to the last full threshold block (main effect of block:  $F(1,22) = 17.71$ ,  $p < .001$ ).

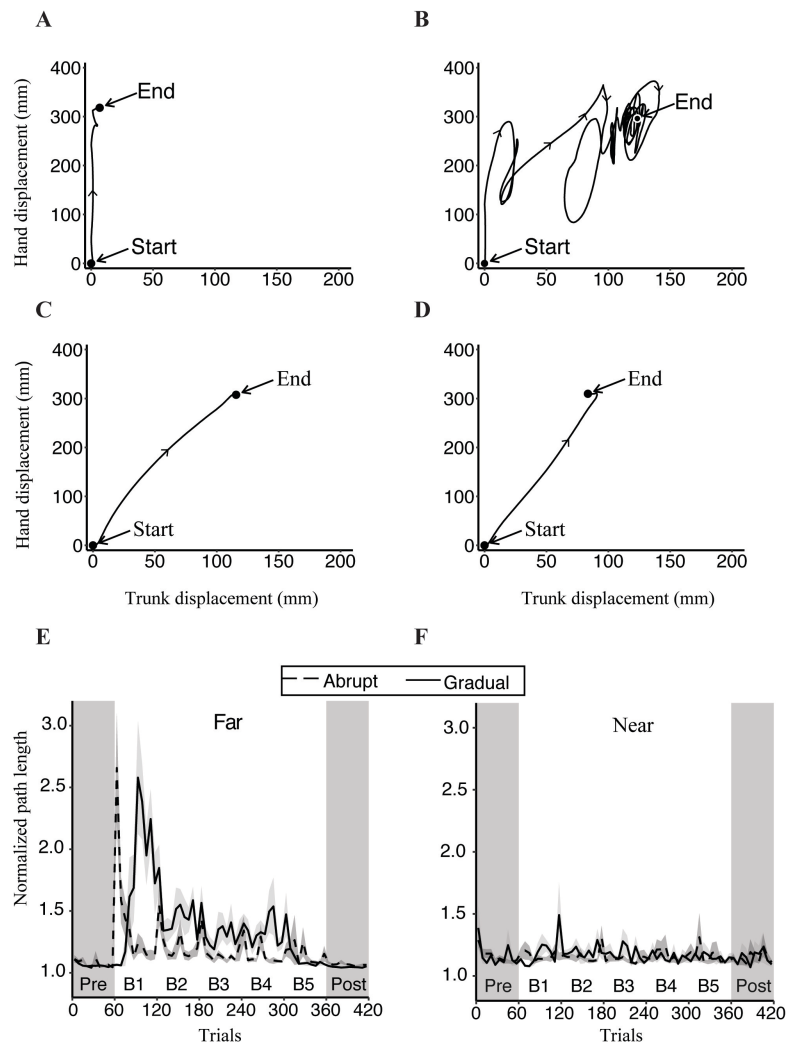
There was no main effect of group ( $F(1,22) = 0.09$ ,  $p = .771$ ), but there was a significant interaction effect between group and block ( $F(1,22) = 4.87$ ,  $p = .038$ ). Post hoc tests showed both

groups had no difference in the first full threshold block ( $p = .200$ ), but that the abrupt group had smaller path length in last full threshold block ( $p = .015$ ).

### ***Path Length - Near targets***

The path length for the near targets is shown in Figure 3.5F.

Adaptation to reinforcement. There was no main effect of block ( $F(1,22) = 2.13$ ,  $p = .159$ ), and no main effect of group ( $F(1,22) = 0.04$ ,  $p = .845$ ), and no interaction effect ( $F(1,22) = 0.01$ ,  $p = .937$ ).



**Figure 3.5. Schematic of path length (A)** Pre-test block – the participant showed a ‘typical’ reaching motion with very little trunk movement. (B) First block with full threshold – participant

showed an exploratory strategy with increase in trunk movement. (C) Last block with full threshold – participant showed lower exploration and simultaneous movement of hand and trunk. (D) Post-test block, without the reinforcement feedback participant still reached the target with more trunk movement compared to the pre-test block. (E) Normalized path length at far targets for the abrupt and gradual groups. (F) Normalized path length at the near targets for the abrupt and gradual groups. Error bars in E and F represent one SEM (between-participant). Both abrupt and gradual groups show initial increases in path length during reinforcement feedback (indicating greater exploration), which decreases with further practice (indicating lesser exploration and simultaneous movement of the hand and trunk).

## **Discussion of Experiment 1 and rationale for Experiment 2**

The results of Experiment 1 showed reinforcement feedback helped to shift coordination in the reaching movement, and that the abrupt group had greater retention of trunk movement during post-test. This was seen both in the trunk-hand distance in the post-test (both far and near targets) as well as the smaller exploration index in the last full threshold block (indicating they had a more coordinated strategy).

We examined two hypotheses that could potentially explain the greater retention of the new movement pattern in the abrupt group in this task. First, we hypothesized that the abrupt group had greater retention because it had more full threshold trials with reinforcement feedback compared to the gradual group (300 trials compared to 180 trials). Secondly, we hypothesized the gradual ‘ramp up’ phase (trials 300-360) of the gradual group resulted in participants becoming aware that the threshold had been removed prior to the post-test, which could have caused them to revert to the typical reaching motion faster.

We added two new groups in Experiment 2 to test these hypotheses - (a) a ‘short abrupt’ group that had the same number of full threshold trials as the gradual group (i.e. 180 trials), and (b) a ‘gradual with abrupt return’ group where the threshold was introduced gradually, but removed abruptly (Figure 3.6D).

## Methods – Experiment 2

All procedures were identical to that mentioned in Experiment 1. We recruited 24 additional healthy college students (Mean age  $\pm$  SD =  $21 \pm 1$  years, 14 Female, 1 left-handed). None of the participants in Experiment 2 were part of Experiment 1. Participants were randomly assigned to one of two groups ( $n=12/\text{group}$ ): a short abrupt group (Figure 3.6A), or a gradual with abrupt return group (Figure 3.6B).

### *Statistical Analysis*

We compared the two new groups to the gradual group separately to show the differences of trunk-hand distance and path length. A group (2) x block (2) mixed design was performed in each of the two comparisons (short abrupt vs. gradual, and gradual with abrupt return vs. gradual). Because the two groups in Experiment 2 were designed in order to examine hypotheses related to the greater retention of the abrupt group in the post-test, we were also only interested in comparing the groups at the post-test (i.e., post-test retention).

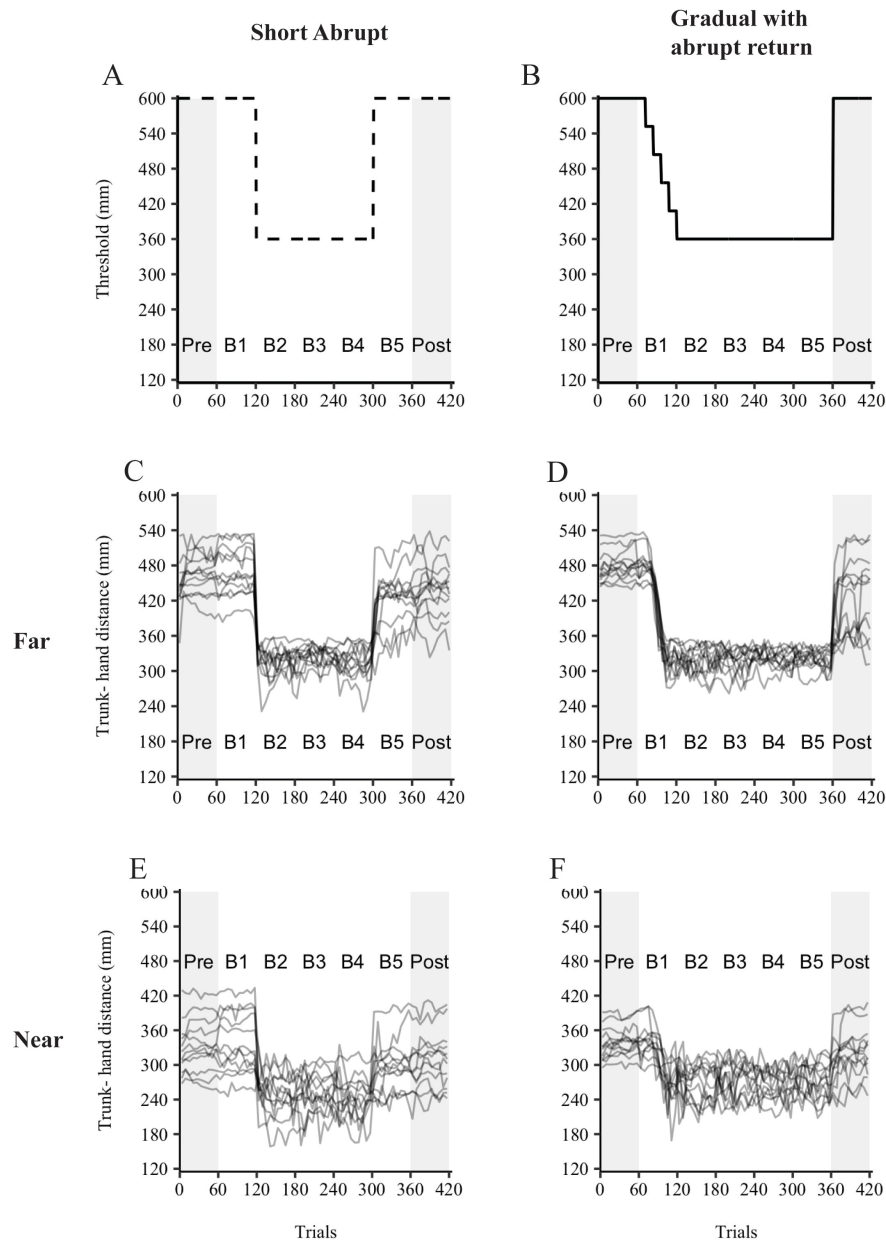
## Results

### *Trunk-hand distance - Far targets*

The trunk-hand distance for the far targets is shown for the short-abrupt (Figure 3.6C) and the gradual with abrupt return group (Figure 3.6D). The average change between the groups is shown in Figure 3.7A.

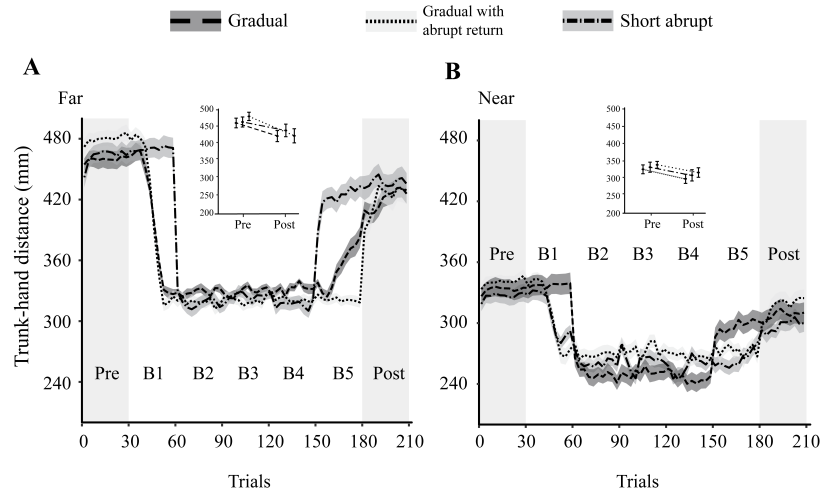
Post-test Retention (Gradual vs. Short abrupt). There was as significant decrease in the trunk-hand distance from pre-test block to post-test block (main effect of block:  $F(1,22) = 15.26$ ,  $p < .001$ ). However, there was no main effect of group ( $F(1,22) = 0.02$ ,  $p = .877$ ), and no Block x Group interaction ( $F(1,22) = 0.01$ ,  $p = .934$ ).

Post-test Retention (Gradual vs. Gradual with abrupt return). There was a significant reduction in the trunk-hand distance from the pre-test block to post-test block (main effect of block:  $F(1,22) = 16.38, p < .001$ ). There was no main effect of group ( $F(1,22) = 0.56, p = .462$ ), and no Block x Group interaction ( $F(1,22) = 0.63, p = .437$ ).



**Figure 3.6. Variation of thresholds and actual trunk-hand distances** (A) Threshold of short abrupt group. (B) Threshold of gradual with abrupt return group. (C) Short abrupt group at far targets. (D) Gradual with abrupt return group at far targets. (E) Short abrupt group at near

targets. (F) Gradual with abrupt return group at near targets. In panels (C-F), each line represents a single participant in that group. Again, similar to Experiment 1, both groups show decreases in trunk-hand distance when reinforcement feedback is provided.



**Figure 3.7. Group mean trunk- hand distance** (A) Mean trunk- hand distance of all participants in the gradual group (from Experiment 1) and the short abrupt group, and gradual with abrupt return groups (from Experiment 2) when reaching to far targets. (B) Mean trunk-hand distance of all participants in the gradual group (from Experiment 1) and the short abrupt group, and gradual with abrupt return groups (from Experiment 2) when reaching to near targets. Error bars represent 1 SEM (between-participant). There was no significant difference between either the gradual and the short-abrupt group, or the gradual and the gradual with abrupt return group in the post-test.

### *Trunk-hand distance - Near targets*

The trunk-hand distance for the near targets is shown for the short-abrupt (Figure 3.6E) and the gradual with abrupt return group (Figure 3.6F). The average change between the groups is shown in Figure 3.7B.

Post-test Retention (Gradual vs. Short abrupt). There was a significant decrease in the trunk-hand distance from pre-test block to post-test block (main effect of block:  $F(1,22) = 9.25$ ,  $p = .005$ ). However, there was no main effect of group ( $F(1,22) = 0.04$ ,  $p = .847$ ), and no Block x Group interaction ( $F(1,22) = 0.11$ ,  $p = .739$ ).

Post-test Retention (Gradual vs. Gradual with abrupt return). There was a significant reduction in the trunk-hand distance from the pre-test block to post-test block (main effect of block:  $F(1,22) = 7.56$ ,  $p = .011$ ). However, there was no main effect of group ( $F(1,22) = 1.80$ ,  $p = .192$ ), and no Block x Group interaction ( $F(1,22) = 0.08$ ,  $p = .781$ ).

## **Discussion of Experiment 2**

The results of Experiment 2 showed that (i) the short abrupt group did not show greater retention compared to the gradual group, and (ii) the gradual with abrupt return group did not show greater retention compared to the gradual group. These results support the idea that rather than the way the threshold was introduced, the number of trials practiced with reinforcement feedback of the desired movement pattern (which was determined by the number of full threshold trials in the current paradigm) determined the retention of the new movement pattern.

## **General Discussion**

We examined the use of reinforcement to shift participants from using one coordination solution to another in a redundant task. Reinforcement in our task was provided by removing vision of the cursor if there was insufficient trunk motion during the reach, as defined by a threshold value. We examined two schedules of introducing the threshold - either abruptly or gradually to examine if there were any differences in retention of the new movement pattern after the threshold was removed.

Regarding our first hypothesis, the results of Experiment 1 showed that both gradual and abrupt schedules of reinforcement were successful in creating changes in coordination from pre-test to post-test. Although the post-test showed signs of a return back to the pre-test movement pattern, it was still significantly different from the movement pattern in the pre-test even after 60

trials of reaching. Regarding our second hypothesis, we found that contrary to our expectation from adaptation studies, the abrupt group showed greater retention of the movement pattern compared to the gradual group. This was seen not only in terms of greater trunk motion (as indicated by the smaller trunk-hand distance), but also a more simultaneous coordinated movement between the trunk and the hand (as indicated by the path length). Experiment 2 further showed that the greater retention of the abrupt group was likely due to the greater number of trials practiced at full threshold, and not due to any specific differences between introducing the threshold in an abrupt or gradual manner. It is important to note that even though we statistically compared two groups from Experiment 2 to a group from Experiment 1 (i.e. they were not prospectively randomized), we are confident in the interpretations of these tests because the participants in these groups were highly similar both from a demographic standpoint, as well as their performance on the pre-test.

These results support the idea of using reinforcement feedback to elicit changes in coordination. Previous studies on reinforcement learning have shown differences in motor adaptation tasks (Izawa & Shadmehr, 2011; Nikooyan & Ahmed, 2015). However as mentioned in the Introduction, these paradigms typically do not generally deal with the issue of learning a novel movement pattern requiring the reorganization of degrees of freedom. In contrast to smooth, exponential changes with practice that is typical of error-driven learning, reinforcement learning resulted in exploration where participants had to ‘break out’ of the typical movement pattern and employ search strategies to discover a new solution. Once this solution was learned, participants still refined the solution by decreasing the path length to produce a more coordinated movement between the trunk and the hand. Our results, along with other recent studies (Mehler et al., 2017; Thorp et al., 2017), show the potential of reinforcement feedback in understanding

learning and coordination in systems with motor redundancy, which could also be extended to rehabilitation of compensatory movements in movement disorders (Michaelson et al., 2006; Ranganathan et al., 2017).

When comparing abrupt and gradual schedules, we found results that are in general counter to the majority of the literature on motor adaptation (Huang & Shadmehr, 2009; Kagerer et al., 1997; Klassen et al., 2005). In our case, the abrupt schedule produced greater retention after the feedback was removed compared to the gradual group. Closer examination of this result in Experiment 2 suggested that the abrupt group had better performance not because of any conscious but because they had greater amounts of specific practice with the required movement pattern (which was determined by the number of full threshold trials). There are two critical differences between the previous studies and the current one. First, the advantage of the gradual group in adaptation studies is attributed to the size of the error signal (Criscimagna-Hemminger et al., 2010). It is hypothesized that smaller errors affect the credit assignment problem, which results in attributing more error to our body than to the environment (Berniker & Kording, 2008). In our study however, we used reinforcement feedback where participants only received binary feedback with no information about the magnitude of error. This difference between the experimental paradigms could explain why the gradual group did not show any advantages, and instead the number of trials at full threshold became the more critical factor. Second, the abrupt group showed a large amount of exploration for a short period of time very early in learning, which was followed by a period of stabilization of the new movement pattern (as indicated by the decrease in path length). In contrast, the gradual group (where the threshold was changed) had to explore by a smaller amount, but over a larger time period. This prolonged exploration meant that they had a smaller time to stabilize the new movement pattern, which potentially

affected retention. Consistent with this hypothesis, when the number of trials was reduced in the abrupt condition in Experiment 2 (for the short abrupt group), there was poorer retention. Further experiments are warranted to fully examine this hypothesis of how gradual vs. abrupt schedules affect stabilization of movement patterns in reinforcement learning.

Finally, we also investigated the generalization of learning the new movement pattern by examining the movements to the near targets. As mentioned earlier, the near targets were positioned close enough that the reinforcement feedback had no influence when reaching to these targets (i.e. the cursor never disappeared when reaching to these targets). Therefore there was no requirement for participants to change their movement pattern when reaching to these targets. Yet, when practicing these near targets in combination with the far targets, participants changed their coordination even for the near targets by using greater trunk motion. These results clearly show that participants were “reusing” the same movement pattern for all targets, even though this is likely more energy inefficient from a metabolic energy standpoint (given the mass of trunk is much larger than the arm). These results support previous studies which found that in tasks requiring the learning of novel movement patterns, the computational cost of changing a solution may be more critical than metabolic cost in determining the movement pattern (Ganesh et al., 2010; Rosenbaum & Jorgensen, 1992).

In summary, we found that reinforcement feedback was capable of causing a change in movement pattern used to perform a redundant task. Although participants moved back closer to the original movement pattern once the feedback was removed (indicating that the effects seen here were more reflective of short-term adaptation than long-term learning), we were able to identify clear changes that persisted over many trials, and across different schedules of training. Understanding how the nervous system organizes and reorganizes movement coordination still

remains a significant challenge in motor learning research and the current results highlight the potential of using reinforcement feedback for changing movement patterns both in healthy and neurologically impaired populations.

## **CHAPTER 4 SHAPING REINFORCEMENT FEEDBACK TO INDUCE CHANGES IN MOVEMENT PATTERNS IN A THROWING TASK**

### **Abstract**

Reinforcement learning has been used to facilitate motor learning, but its applicability in multiple degree of freedom (DOF) tasks with redundancy is not fully understood. A critical issue that arises in such tasks is how to use reinforcement to guide exploration in a high-dimensional space toward the desired movement patterns. Here, using two experiments, we examined the use of reinforcement feedback and different shaping techniques to change movement patterns in a multi-DOF task with redundancy. 70 college-aged participants performed a virtual throwing task where the goal was to throw a ball toward a target, and we introduced redundancy by making the ball velocity a linear combination of trunk and hand velocity. The goal of the participants was to use reinforcement feedback, provided as a score after each trial, to shift their movement pattern toward a higher trunk velocity. We used different shaping techniques to manipulate the threshold of trunk velocity below which participants received ‘punishment’ (i.e. a bad score) and examined the change in the trunk velocity with practice. In experiment 1, we compared three shaping techniques (abrupt, gradual and adaptive) that manipulated the threshold in different ways. In experiment 2, we compared four adaptive shaping techniques to further examine what characteristics of adaptive feedback. We found that: (i) reinforcement feedback in multi-DOF task was less successful in changing movement patterns compared to previous studies using single DOF tasks, (ii) adaptive shaping techniques, that incorporated the participant’s current level of performance, were more successful in changing movement patterns. These results highlight the potential of adaptive shaping techniques for learning complex motor skills.

## Introduction

Motor redundancy (or alternatively motor abundance (Latash, 2012)) , which arises due to the large number of degrees of freedom (DOF) in the human body, allows humans to perform most motor tasks with multiple movement patterns (Bernstein, 1967). This redundancy is present at several levels of analysis (trajectories, joints, muscles etc.), and understanding the organization of these DOFs with practice has been one of the central questions in motor learning (Martin et al., 2011; Müller & Sternad, 2004; Newell, 1986; Newell et al., 2003; Newell & Vaillancourt, 2001; J. P. Scholz & Schöner, 1999; Vereijken et al., 1992; Yang & Scholz, 2005).

The phenomenon of motor redundancy raises an important issue from a motor learning standpoint. Specifically, although most studies of motor learning focus on a change in the task performance (e.g., increasing speed or decreasing variability) the fact that movement patterns and task outcomes do not have a one-to-one relation means that motor learning could involve changes in movement pattern *without* creating associated changes in the movement outcome (Latash et al., 2002). This aspect of motor learning is especially critical in situations when the movement pattern (rather than task performance) is the primary target of the intervention. For example, a golfer may want to alter their swing to reduce the risk of injury even if it does not result in any benefits in terms of their golf score. Similarly, in rehabilitation, stroke survivors are often guided away from the use of atypical compensatory movement patterns because they may impact long-term rehabilitation (Levin et al., 2009). Thus, the question of how to effectively create such changes in movement patterns in complex tasks with multi-DOF movements is still poorly understood.

One of the challenges in using standard ‘prescriptive’ motor learning techniques (such as feedback and instructions etc.) to create changes in multi-DOF movements is the problem of

which information to specify. Because complex tasks involve the coordination of several DOFs, it is a challenge to get the learner to understand how to change each DOF simultaneously. For example, even a simple reaching motion involves motion at 7 DOFs (3 at the shoulder, 2 at the wrist, 1 at the elbow and 1 at the forearm). Therefore, trying to provide precise feedback or instructions on how to change each of these 7 DOFs after a given trial is likely not effective because it creates an ‘information overload’ for the learner (Wulf & Weigelt, 1997). Other techniques, such as visual demonstrations, help participants get a general idea of the intended movement pattern, but again, correcting specific errors at each DOF is difficult with such approaches.

A potential solution to get around this problem is through reinforcement feedback. Reinforcement feedback is a signal that evaluates the whole performance then provides a simplified signal (either a binary or a scalar) to the learners (Wolpert et al., 2001). Because the feedback is not provided at the level of each DOF, participants typically engage in ‘exploration’ to learn alternative movement patterns (Dhawale et al., 2017). By exploring different movement patterns, participants could eventually learn the desired movement pattern that returns good feedback. Several studies of motor learning have shown that reinforcement feedback can be used to effectively modify movement patterns (Chen et al., 2018; Gläscher et al., 2010; Izawa & Shadmehr, 2011; Therrien et al., 2016).

However, a key limitation of these studies is that they used simple tasks with limited or no redundancy. In redundant tasks with multiple DOFs, the exploration process is more complex because of the need to explore a higher dimensional space. Moreover, because these tasks require coordination of multiple degrees of freedom, there may be a greater possibility for ‘preferred’ coordination patterns that are more resistant to change (Schoner & Kelso, 1988). Therefore, there

is a need to examine ‘adaptive shaping techniques’ that can be used to guide exploration more effectively. Prior work has shown that adaptive reinforcement can be successfully used in modifying movement patterns in low dimensional tasks (Therrien et al., 2016). Here, we examine the use of different shaping techniques (Skinner, 1938) to change movement patterns in multi-DOF tasks.

In this study, we performed two experiments where we used reinforcement feedback to guide participants to change to an alternative movement pattern in a multi-DOF task. The task was a virtual throwing task requiring coordination of the trunk and hand velocities. Typically, participants perform this task primarily with the hand (i.e. large hand velocity and small trunk velocity) – so our goal was to examine if we could get participants to increase their trunk velocity when throwing the ball by providing reinforcement feedback based on setting a threshold for the trunk velocity. We then compared the change in movement patterns between different shaping techniques that modified how this threshold across practice. In experiment 1, we compared three shaping techniques (abrupt, gradual and adaptive). While the abrupt and gradual groups changed the threshold in an ‘open-loop’ manner (i.e. the same way regardless of the participant performance), the adaptive schedule changed the threshold based on the participants’ prior performance (Therrien et al., 2016). In experiment 2, we compared four adaptive shaping techniques to further examine what characteristics of an adaptive schedule were most effective in facilitating a greater change of movement patterns.

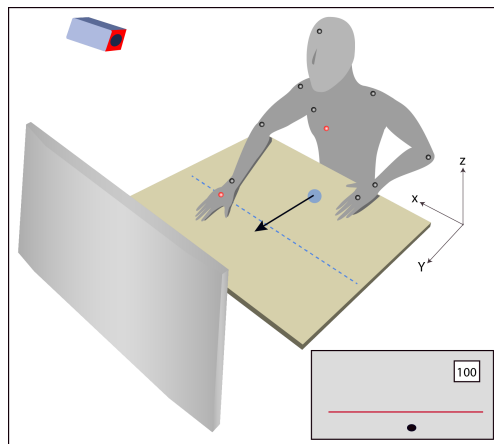
## Methods: experiment 1

### *Participants*

Thirty healthy college students (age range: 18-21 years, 20 females) with no upper body injuries were recruited to participate in the experiment for extra course credit. All participants provided informed written consent and all procedures were approved by the Institutional Review Board at Michigan State University.

### *Apparatus*

Participants sat comfortably with both arms resting on the desk and faced a 50" (125 cm) television screen. The virtual throwing task was implemented with a 120 Hz Motion Capture system (Motion Analysis Corporation, Santa Rosa, CA), and a MATLAB program. Participants wore 11 retro-reflective markers (forehead, sternum, right chest, and bilaterally at the shoulder, elbow, wrist, and the third metatarsophalangeal joint) (Fig. 4.1)



**Figure 4.1. Experimental setup.** Participants wore 11 retro-reflective markers. The ball on the screen was controlled by the right hand marker. The task was to glide the right hand through the line and maximize the score. There was no visual feedback on the screen except the score.

### ***Task***

Participants were instructed to glide their right hand on the desk to throw a virtual ball to a target. The MATLAB program received the (x, y, z) coordinates of the retro-reflective markers from the motion capture system and mapped the (x, y) coordinates (corresponding to the plane of the desk) of the right metatarsophalangeal marker to a ball on the screen. Participants controlled the ball by the right hand and glided the right hand through a line on the screen. When the right hand passed this line, the ball was ‘released’ and the program computed the instantaneous velocity of y-direction (moving forward toward the screen) of the right metatarsophalangeal marker (to approximate hand velocity) and the sternum marker (to approximate trunk velocity). The ball velocity was then calculated based on the following equation:

$$V_{ball} = 4 * V_{trunk} + V_{hand}$$

The goal for the participants was to throw the ball with a velocity of 1100 mm/s. This task is redundant because the same ball velocity can be achieved with different hand and trunk velocities. The trunk velocity was scaled up by 4 times based on pilot testing to make the variation in trunk velocity comparable to the variation in the hand velocity.

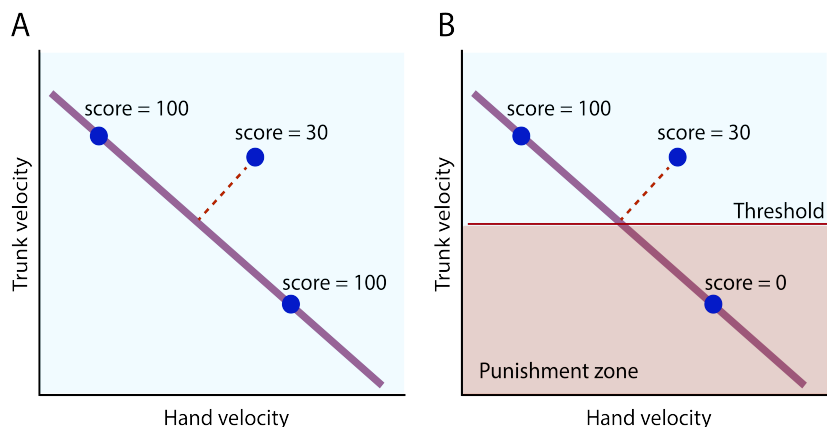
### ***Score feedback***

Participants did not see the target displayed on the screen, instead the only feedback they received after each throw was a score that reflected their performance. The score was calculated based on the following equation (Figure 4.2A):

$$Score = 100 - \frac{90}{(1 + 5e^{-0.014x})^5}$$

Where x is the absolute error in the ball velocity (i.e. the absolute difference between the actual ball velocity and the target velocity of 1100 mm/s)

The maximum score was 100 and the minimum was 10. Participants were not informed about how the score was computed. They were only instructed to try different movement patterns to maximize the score.



**Figure 4.2. Mechanism of providing reinforcement feedback** (A) Each blue point represents the combination of trunk and hand velocity. The Euclidean distance between the point and the solution manifold was mapped to the score function to calculate the score. The minimum score is 0, and the maximum score is 100. Different points could have the same score. (B) Reinforcement feedback was provided by adding a punishment zone to the state space. Participants received 0 score in the punishment zone even they were on the solution manifold. The size of the punishment zone was set by the trunk velocity threshold.

### *Providing reinforcement feedback*

The goal was to examine how reinforcement feedback could be used to shift participants to use a higher trunk velocity when throwing the ball. The desired trunk velocity was set at 100 mm/s based on pilot testing – this number was high enough that participants could perceive this change as a different coordination pattern, but not so high to be uncomfortable for participants. Reinforcement was provided based on a trunk velocity threshold (Fig. 4.2B) and how this threshold was changed with practice differed based on the group participants were assigned to. When the trunk velocity was larger than the threshold, the score was calculated from the equation. When the trunk velocity was smaller than the threshold, the score was zero. It is important to note that because the score also reflected their performance on the task (i.e. how

close the throwing velocity was close to 1100 m/s), we used the “0-score” to signal that participants that they did something incorrectly (the minimum score otherwise was 10). The reinforcement feedback acted as punishment – i.e., we expected participants to increase the trunk velocity to avoid the 0 scores (Figure 4.2B).

### ***Procedures***

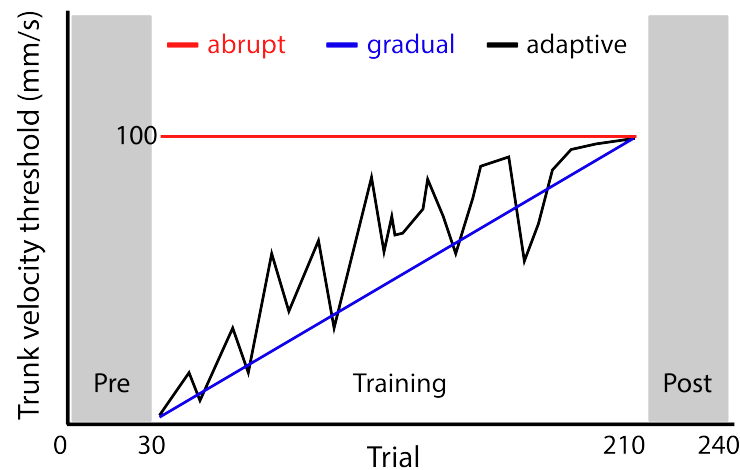
Each participant performed 8 blocks of 30 throws for a total of 240 trials in the experiment. Participants started from a pre-test block without reinforcement feedback (30 trials). Subsequently, they practiced 6 training blocks with reinforcement feedback ( $6 \times 30 = 180$  trials). Lastly, they performed a post-test block without reinforcement feedback (30 trials). In the pre-test and post-test, participants received normal score feedback. In the training blocks, we added one additional constraint to the score function to construct the reinforcement feedback. Participants were not informed about the onset and offset of the reinforcement feedback.

### ***Groups***

Reinforcement feedback was introduced in the training blocks with three shaping methods that varied in how the threshold was set for the trunk velocity across practice: abrupt, gradual, and adaptive.

In the abrupt group, the threshold was increased to 100 mm/s in the first training block and remained constant until the last training block. In the gradual group, the threshold started from 0 mm/s in the first training block and linearly increased to 100 mm/s in the last training block. In the adaptive group, the threshold was calculated from the past performance of the participant. We took the average trunk velocity of the previous 6 trials to set the current threshold. Furthermore, to avoid overshooting the desired trunk velocity, the program set the

threshold as 100 mm/s when the average trunk velocity of the previous 6 trials exceeded 100 mm/s (Figure 4.3).



**Figure 4.3. Design of shaping methods.** Abrupt group had huge rate of change of the threshold after pre-test, the threshold was fixed in the rest of the training. Gradual group had a positive linear change trial-by-trial in the training. The rate of change of threshold in Adaptive group was set by the average of previous 6 trials of the participant. Each participant in the Adaptive group had their own unique schedule depending on their performance.

### *Data analysis*

#### Score

The score feedback at the end of each trial was our measure of task performance, with higher scores indicating better task performance (i.e. the ball velocity was closer to 1100 m/s). In pre-test and post-test, there was no reinforcement feedback (max score was 100 and min score was 10). In the training blocks (B1 to B6), when the participants received reinforcement feedback, the score was 0, when participants did not receive reinforcement, the maximum score was 100, and the minimum was 10.

#### Punishment rate per block

The punishment rate in a block was the number of total trials where the participant was punished (i.e. received a score of zero because the trunk velocity was under the threshold) divided by the total number of trials in that block. A reduction in the punishment rate indicated

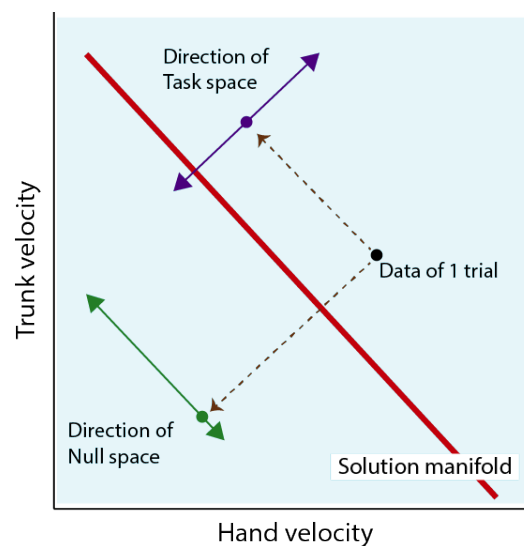
that participants increased their trunk velocity above the threshold. Because the pre-test and post-test did not have reinforcement, the punishment rate was computed only for the training blocks.

### Trunk and hand velocity

The trunk and hand velocity were measured at the instant where the ball was released. Because the trunk velocity was multiplied by four to compute ball velocity, we report this value as the “standardized trunk velocity”. All the trunk velocities in the result section are standardized trunk velocity.

### Task and null space variability

The exploration during the task was captured using the movement variability. In a redundant system, the movement variability can be separated into task space variability and null space variability (Latash et al., 2002; Mussa-Ivaldi et al., 2011; Ranganathan et al., 2014; Scholz & Schöner, 1999). The task space is the direction that is orthogonal to the solution space, where variability is detrimental to task performance. On the other hand, the null space is along the direction of the solution space, where variability does not affect task performance (Figure 4.4).



**Figure 4.4. Definition of task space and null space.** Task space is the purple arrow, it is orthogonal to the solution manifold. Null space is the green arrow, it is parallel to the solution

manifold. To calculate the variability, the data were projected to both spaces to calculate the variance of all the projected points on each space.

### ***Statistical analysis***

We had two main research questions of interest – (i) in the training phase, how did reinforcement feedback affect movement for the different groups, and (ii) in the test phase, was the effect of reinforcement feedback present after the feedback was removed? To analyze this, we separated our analysis on the training phase and test phases. We used R (3.5.1) to run RM-ANOVA for all the dependent variables, and the alpha value was set as .05.

#### Training phase

First, we compared the change between the first block with reinforcement feedback (B1) and the last block with reinforcement feedback (B6) to investigate how participants reacted to the reinforcement feedback. We ran a block (2) X group (3) RM-ANOVA with block as the within-subject factor, and the group as the between-subject factor.

#### Test phase

Second, we compared the change between pre-test and post-test to investigate the overall change of practicing with different shaping methods. We ran a block (2) X group (3) RM-ANOVA with block as the within-subject factor, and the group as the between-subject factor.

### **Results: experiment 1**

To screen for outliers at the trial level, we used the Mahalanobis Distance (De Maesschalck et al., 2000). Approximately 3% of trials were removed based on this criterion before running the statistical tests. At the participant level, we also screened for high trunk velocities at the pre-test using a boxplot (because the goal of the study was to examine an

increase in trunk velocity). One participant in abrupt group and one participant from adaptive group were removed based on this criterion.

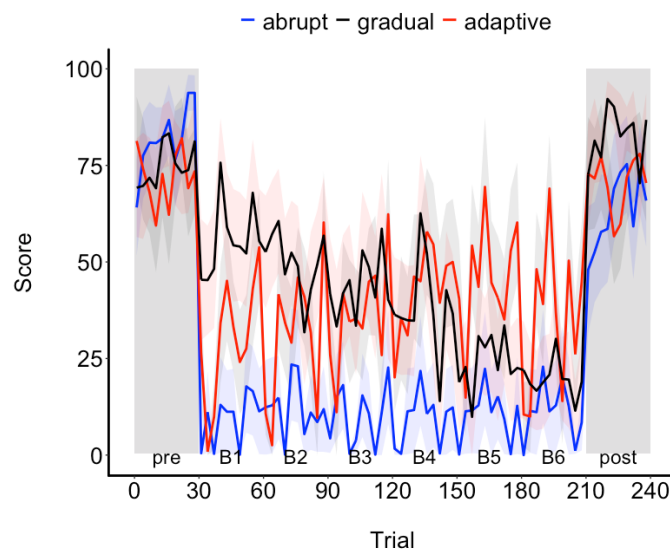
## ***Score***

### Training phase

There was a significant main effect by group ( $F(2,25) = 8.35, p = .002$ ), and group x block interaction ( $F(2,25) = 7.03, p = .004$ ). The main effect by block was not significant ( $F(1,25) = 4.0, p = .06$ ). Analysis of the group x block interaction revealed that the gradual group decreased the scores from B1 to B6; the change in other two groups was not significant (Figure 4.5).

### Test phase

There was no significant main effect by block ( $F(1,25) = .28, p = .76$ ), group ( $F(2, 25) = 0.14, p = .71$ ) or group x block interaction ( $F(2, 25) = 1.70, p = .20$ ) (Figure 4.5).

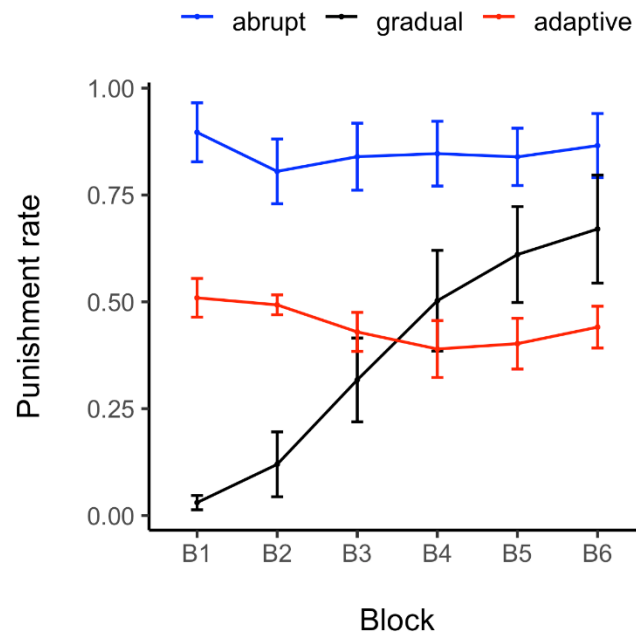


**Figure 4.5. Change in mean score.** The score dropped dramatically after reinforcement feedback was introduced because participants received punishment based on their trunk velocity. There was no reinforcement in the post-test so the average score increased back to the level of pre-test. Error bars indicate 1 SE (between-participants).

## ***Punishment rate***

### Training phase

There was a significant main effects by block ( $F(1,25) = 10.27, p=.004$ ), group ( $F(2,25) = 25.07, p < .001$ ), and a group x block interaction ( $F(2, 25) = 17.37, p < .001$ ). The analysis of the interaction revealed that in block B1, all three groups were significantly different from each other (with abrupt having the highest punishment rate, and gradual the lowest punishment rate). In block B6, the punishment rate in the abrupt group was higher than the adaptive group (Figure 4.6).



**Figure 4.6. Change in mean punishment rate.** Abrupt group had constant high punishment rate throughout the experiment. Gradual group did not receive punishment at the beginning then gradually received more punishment toward the end of the experiment. Adaptive group had constant moderate punishment rate throughout the experiment.

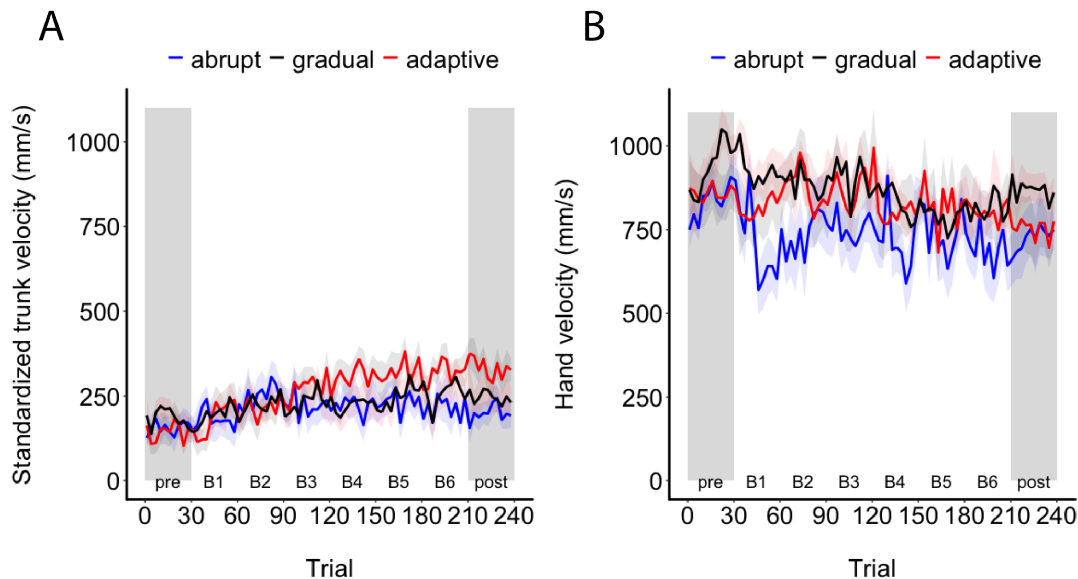
## ***Trunk velocity***

### Training phase

There was a significant main effect by block ( $F(1, 25) = 10.58, p = .003$ ), with higher trunk velocities in B6 compared to B1. There was no significant main effect by group ( $F(2, 25)$ ,  $p = .44$ ) or group x block interaction ( $F(2, 25) = 0.73, p = .49$ ) (Figure 4.7A).

#### Test phase

There was a significant main effect by block ( $F(1, 25) = 23.47, p < .001$ ), with higher trunk velocity in post-test compared to pre-test. There was no significant main effect by group ( $F(2, 25) = 1, p = .38$ ). There was a significant group x block interaction effect ( $F(2, 25) = 5.80, p = .009$ ). Trunk velocities increased from pre- to post-test but the interaction revealed that in the pre-test, there were no significant differences between groups; however in the post-test, the adaptive group had higher trunk velocities compared to the abrupt and gradual groups. (Figure 4.7A).



**Figure 4.7. Change in standardized trunk velocity and hand velocity** (A) Compare pre-test and post-test standardized trunk velocity had the trend to go up in all three groups. Adaptive group had the largest increment in the post-test. (B) Hand velocity did not change much between pre-test and post-test.

#### *Hand velocity*

#### Training phase

There was no significant main effect by group ( $F(2, 25) = 1.74, p = .20$ ), no significant main effect by block ( $F(1, 25) = 0.87, p = .36$ ), and no significant group x block interaction effect ( $F(2, 25) = 1.48, p = .25$ ) (Figure 4.7B).

#### Test phase

There was a significant main effect by block ( $F(1, 25) = 4.55, p = .04$ ), with a decrease in hand velocity in the post-test relative to the pre-test. There was no significant main effect by group ( $F(2, 25) = 1.96, p = .16$ ) and no significant group x block interaction ( $F(2, 25) = 0.02, p = .98$ ) (Figure 4.7B).

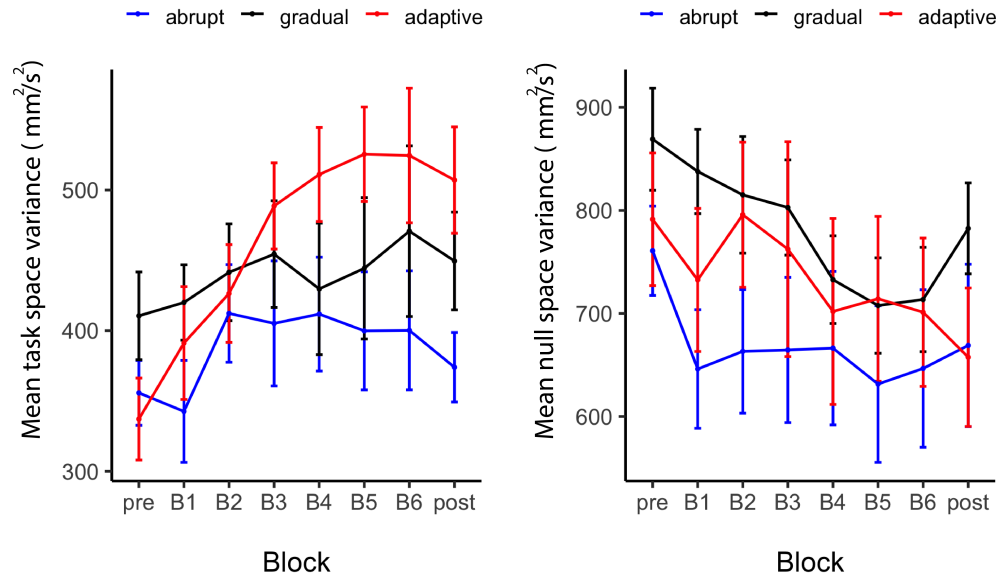
#### ***Task space variability***

##### Training phase

There was a significant main effect by block ( $F(1, 25) = 7.32, p = .01$ ), the task space variability increased in B6 compared to B1. There was no significant main effect by group ( $F(2, 25) = 1.68, p = .21$ ) and no significant group x block interaction ( $F(2, 25) = 0.78, p = .47$ ) (Figure 4.8B).

##### Test phase

There was a significant main effect by block ( $F(1, 25) = 12.95, p = .001$ ), the task space variability increased in B6 compared to B1. There was no significant main effect by group ( $F(2, 25) = 2.02, p = .15$ ). There was a significant group x block interaction effect ( $F(2, 25) = 4.98, p = .002$ ). The interaction showed there was no difference in the pre-test, but in the post-test, adaptive group had higher task space variability than the other two groups (Figure 4.8B).



**Figure 4.8. Change in task and null space variability** (A) Task space variability of abrupt and gradual group remained similar throughout the experiment. Task space variability increased in adaptive group. (B) Null space variability decreased in all three groups.

### *Null space variability*

#### Training phase

There was no significant main effect by group ( $F(2, 25) = 1.39$ ,  $p = .27$ ), no significant main effect by block ( $F(1, 25) = 2.80$ ,  $p = .11$ ), and no significant interaction effect ( $F(2, 25) = 1.51$ ,  $p = .24$ ) (Figure 4.8B).

#### Test phase

There was a significant main effect by block ( $F(1, 25) = 7.32$ ,  $p = .01$ ), null space variability was lower in the post-test compared to pre-test. There was no significant main effect by group and no significant main effect by group ( $F(2, 25) = 1.68$ ,  $p = .21$ ) group x block interaction ( $F(2, 25) = 0.78$ ,  $p = .47$ ) (Figure 4.8B).

### ***Summary of experiment 1***

We investigated how to use reinforcement feedback with shaping to guide participants to change their movement pattern in a redundant task. Participants performed a virtual throwing task involving the coordination of the hand and trunk where reinforcement was provided to increase trunk velocity using different schedules (abrupt, gradual and adaptive). The main results showed that (i) exploration in all schedules were somewhat suboptimal as reflected by the high punishment rates at the end of learning (upwards of 50%), and (ii) although all groups increased their trunk velocity with practice, the adaptive group had the greatest change in trunk velocity from the pre- to the post-test.

### ***Rationale for Experiment 2***

Our results found that the adaptive group had the greatest change in trunk velocity from pre- to post- test. Our adaptive algorithm was based on a simple average of the previous 6 trials – so we examined if modifying the parameters of the adaptive algorithm would have further benefits. We considered two parameters – memory (i.e. how many trials the adaptive algorithm acts over), and momentum (whether the adaptive algorithm uses the trend in those trials). We anticipated that this would provide further understanding regarding why the adaptive algorithm was more successful.

### **Methods- experiment 2**

Experiment 2 had four groups and shared the same procedure and data analysis with experiment 1. None of the participants in Experiment 2 had participated in Experiment 1. Given the relatively poor response to reinforcement in Experiment 1, we made two modifications to the experimental procedure in Experiment 2 - the score feedback and the instructions to the

participant. The rationale for modifying the score feedback was as follows: in experiment 1, the lowest score available without reinforcement was 10 while trials that received punishment got a score of zero. Given that punishment rates did not decrease with practice, we thought that this difference of “10 points” might have been too small for participants to perceive as punishment. So in Experiment 2, we increased this difference to be 200 points (with a maximum score of 1000 points). The rationale for changing the instructions was to examine if the exploration in Experiment 1 was suboptimal because participants did not know what body segments contributed to the task performance. As a result, we instructed the participants to focus on movements of the right hand and the trunk to examine if this would improve the exploration.

### ***Participants***

Forty healthy college students (age range: 18-24, 28 females, 4 left-handed) were recruited to participate in the experiment for earning extra credits. Left-handed participants also performed the task with their right hand. Because the focus of the task was only on generating a given velocity at release (and not trajectory or steady-state control), we did not expect handedness to play a major role. All participants signed the consent form and the consent process was approved by the Institutional Review Board at Michigan State University.

### ***Apparatus and task***

The experimental setting of experiment 2 was identical to experiment 1, with only the score feedback modified as shown below:

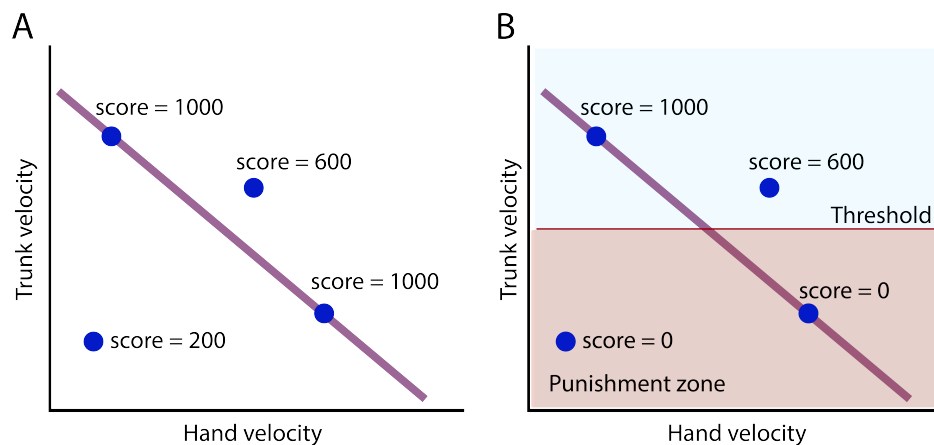
$$V_{ball} = 4 * V_{trunk} + V_{hand}$$

$$Score = 1000 - \frac{800}{(1 + 5e^{-0.014x})^5}$$

Where x is the absolute difference of ball velocity and target velocity (=1100 mm/s)

The maximum score is 1000 and the minimum score is 200. (Figure 4.9A).

During the training blocks, the reinforcement feedback was provided as the punishment of a score of zero. When the trunk velocity was smaller than the threshold, the participants received a zero. When the trunk velocity is larger than the threshold, the score was calculated by the equation where the maximum is 1000 and the minimum is 200. This made the zero score a punishment indicating something needed to be corrected.

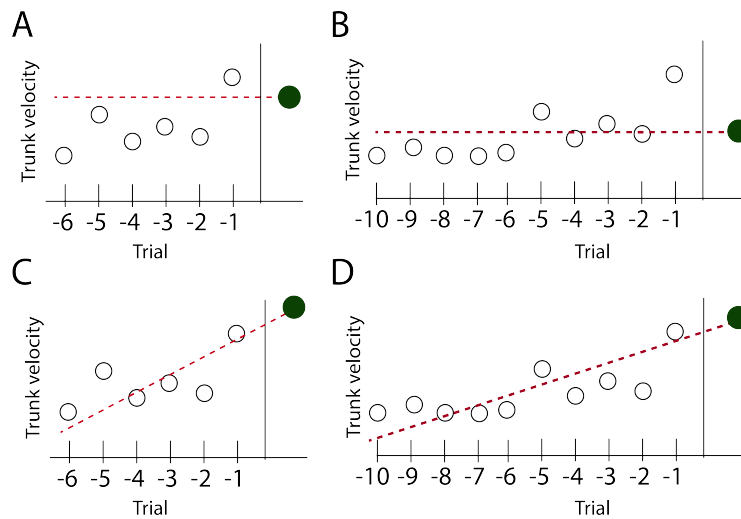


**Figure 4.9. Schematic of scoring with and without reinforcement feedback** (A) The score is 1000 when on the solution manifold (the purple line). When off the manifold the score was calculated by the distance between the point and the manifold, closer to the line the score is higher. (B) Adding a punishment zone to the state space. The punishment zone was defined by the trunk velocity. When the trunk velocity was smaller than then threshold the score would be 0 no matter how far to the solution manifold.

### ***Grouping and shaping methods***

We manipulated two parameters, memory and momentum in a crossed fashion, resulting in 4 groups. The memory parameter was changed by controlling the number of trials involved in the moving average. The momentum parameter was changed by using (or ignoring) the trend of how the velocities changed in that time window. This resulted in the following groups: 1) adaptive 2) long-adaptive 3) momentum 4) long-momentum. The adaptive group was the same as in the previous study, where a moving average of six prior trials was used to calculate the threshold of the next trial. In the Long-adaptive group, we increased the window size to ten prior

trials. In both groups, the threshold was computed using a moving average meaning that no trends in the data were used. In the Momentum group, we aimed to increase the speed of adjusting the threshold by using a linear regression over the prior six trials to predict the threshold on the next trial. Finally, in the Long-momentum group, we used the same strategy as the Momentum group but used a linear regression that included the previous 10 trials to predict the next threshold (Figure 4.10).



**Figure 4.10. Design of reinforcement schedules** (A) In the adaptive group in experiment 1, the threshold was the average of the previous six trials. (B) Long-adaptive group, increasing the window size (memory) of the average to 10, this shows how participants use delayed feedback to adjust motor behavior. (C) Momentum group, use a function to predict the threshold, so we can keep the momentum and create a smooth change. (D) The long-momentum group manipulates both window size and momentum

### *Data analysis*

We kept the same dependent variables and ran the same statistical analyses as in experiment 1. Additionally, we showed how the threshold changed in different shaping methods.

## Results: experiment 2

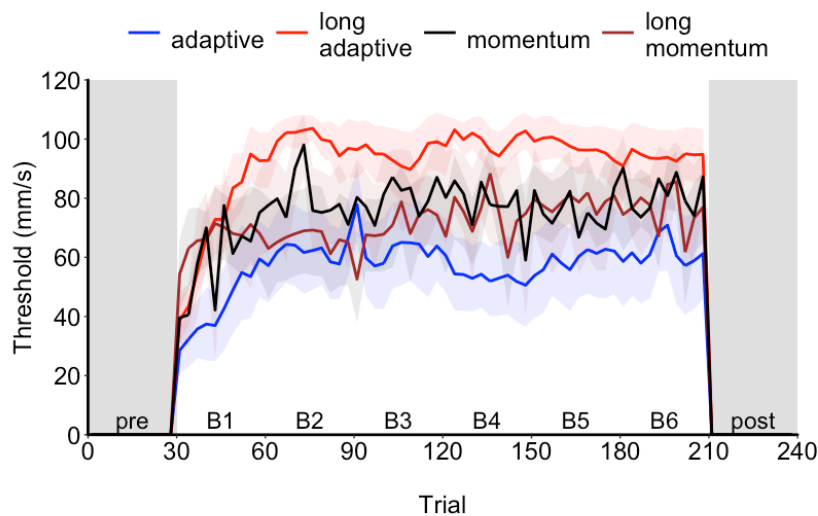
Following the same procedure as in experiment 1, we removed the participants who had high trunk velocity in the pre-test. There was 1 participant removed from each group. Then we used Mahalanobis Distance to remove the outliers before running the statistical test (De Maesschalck, Jouan-Rimbaud, & Massart, 2000). There were around the 5 % of trials got removed.

### *Threshold*

#### Training phase

All four groups increased the threshold in the training phase. There was a significant main effect by block ( $F(1, 32) = 11.68, p = .002$ ), indicating that the threshold increased from B1 to B6.

There was no significant main effect by group ( $F(3, 32) = 1.87, p = .15$ ) and no significant group x block interaction effect ( $F(3, 22) = 0.36, p = .78$ ) (Figure 4.11).



**Figure 4.11. Change in trunk velocity threshold.** All the groups have higher threshold at the last training block but there was no significant group difference.

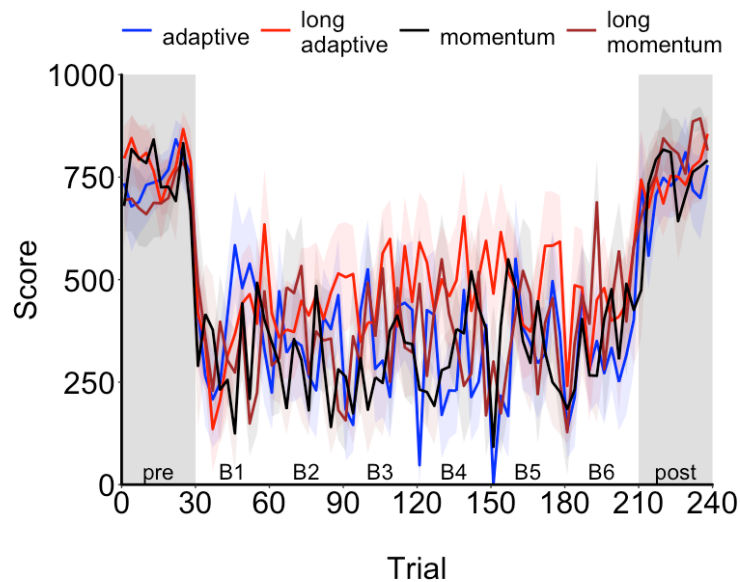
### *Score*

#### Training phase

There was no significant main effect by block ( $F(1,32) = 2.61, p = .12$ ), there was no significant main effect by group ( $F(3, 32) = 0.41, p = .75$ ), and no significant group x block interaction effect ( $F(3,32) = 0.61, p = .61$ ) (Figure 4.12).

#### Test phase

There was no significant main effect by block ( $F(1,32) = 0.69, p = .41$ ), no significant main effect by group ( $F(3,32) = 0.18, p = .91$ ), and no significant group x block interaction effect ( $F(3,32) = 0.48, p = .70$ ) (Figure 4.12).

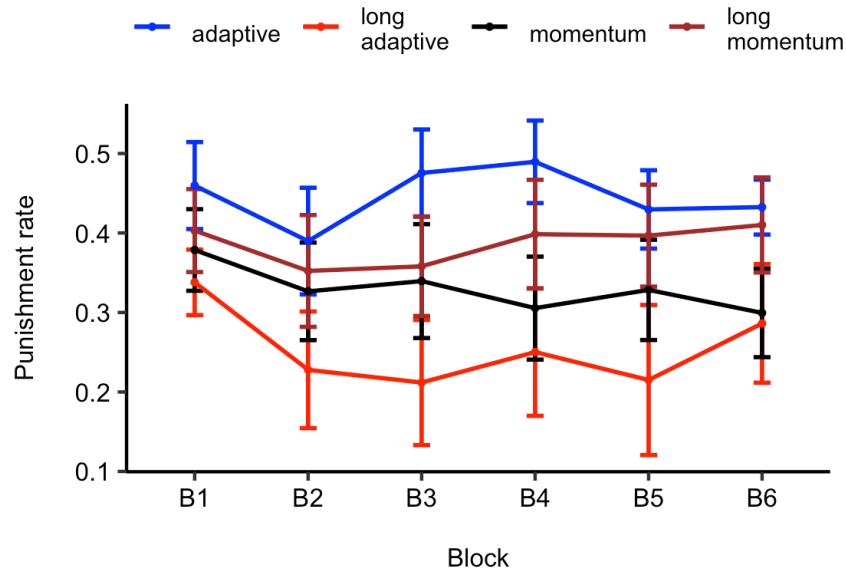


**Figure 4.12. Change in mean score.** In the pre-test and post-test, there was no difference between the four groups.

#### ***Punishment rate***

#### Training phase

There was no significant main effect by block ( $F(1,32) = 1.50, p = .23$ ), there was no significant main effect by group ( $F(3, 32) = 1.90, p = .15$ ), and no significant group x block interaction ( $F(1,32) = 0.69, p = .41$ ) (Figure 4.13).



**Figure 4.13. Change in punishment rate.** Adaptive group (blue) had the highest overall punishment rate. Long adaptive group (red) had the lowest overall punishment rate. Momentum group (black) and long momentum (brown) were in between where long momentum had higher average.

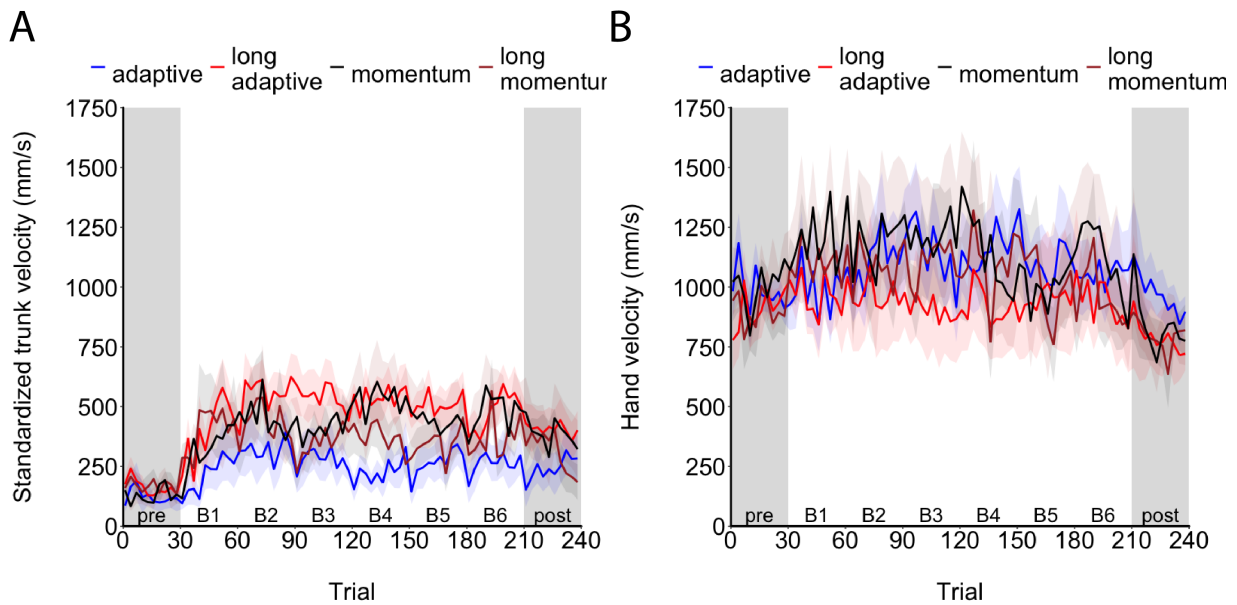
### *Trunk velocity*

#### Training phase

There was no significant main effect by block ( $F(1,32) = 2.50, p = .07$ ), no significant main effect by group ( $F(3, 32) = 2.08, p = .12$ ) and no significant group x block interaction ( $F(3, 32) = 0.72, p = .67$ ) (Figure 4.14A).

#### Test phase

There was a significant main effect by block ( $F(1,32) = 18.74, p < 0.001$ ), trunk velocity was higher in the post-test. There was no significant main effect by group ( $F(3,32) = 0.70, p = .56$ ) and no significant group x block interaction effect ( $F(3,32)=0.91, p = .45$ ) (Figure 4.14A).



**Figure 4.14. Change in standardized trunk velocity and hand velocity** (A) Mean standardized trunk velocity over trials. After pre-test all four groups increased the trunk velocity. (B) Mean hand velocity over trials. There was no trend in hand velocity.

### *Hand velocity*

#### Training phase

There was no significant main effect by block ( $F(1,32) = 0.41, p = .53$ ), no significant main effect by group ( $F(3, 32) = 0.37, p = .78$ ), and no significant group x block interaction ( $F(3, 32) = 0.37, p = 0.78$ ) (Figure 4.14B).

#### Test phase

There was no significant main effect by block ( $F(1,32) = 4.74, p = .04$ ), no significant main effect by group ( $F(3, 32) = 1.28, p = .30$ ), and no significant group x block interaction ( $F(3, 32) = 0.21, p = .89$ ) (Figure 4.14B).

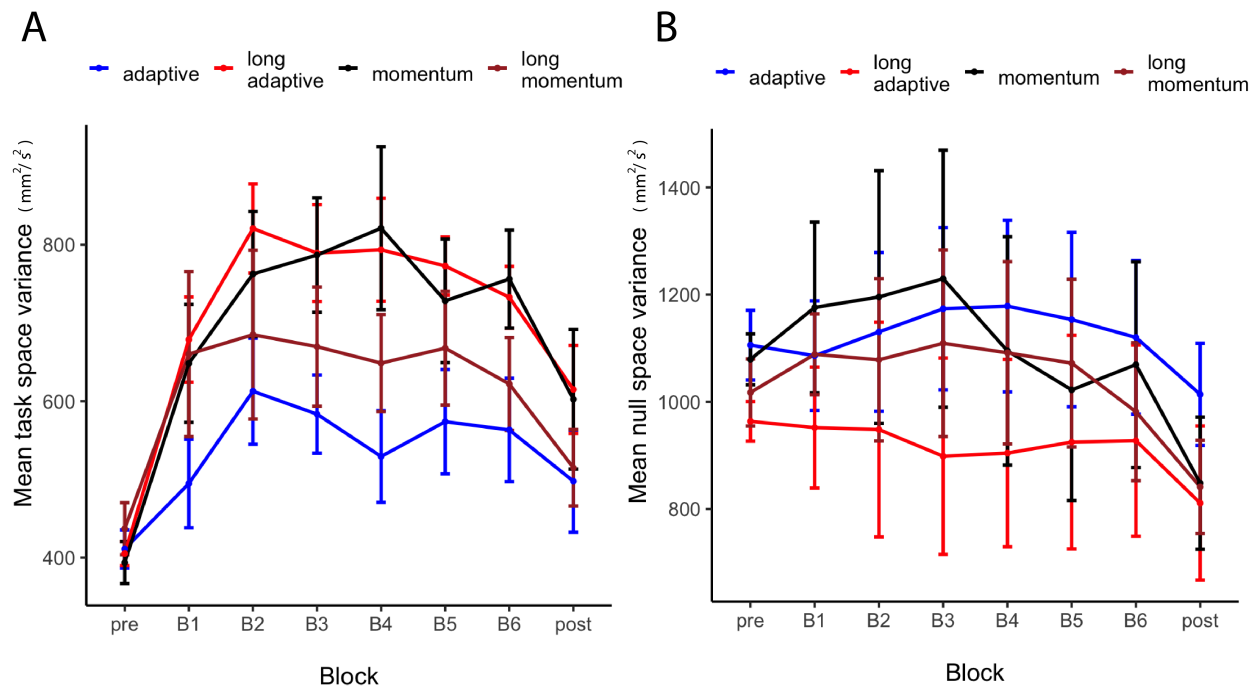
### *F. Task space variability*

#### Training phase

There was no significant main effect by block ( $F(1,32) = 1.55, p = .22$ ), no significant main effect by group ( $F(3, 32) = 2.23, p = .10$ ), and no significant group x block interaction ( $F(3, 32) = 0.64, p = .60$ ) (Figure 4.15A).

### Test phase

There was a significant main effect by block ( $F(3, 32) = 19.43, p < .001$ ), task space variability was higher in the post-test compared to pre-test. There was no significant main effect by group ( $F(3,32) = 0.41, p = .74$ ) and no significant group x block interaction ( $F(3,32)=1.23, p = .31$ ) (Figure 4.15A).



**Figure 4.15. Change in task and null space variability** (A) Task space variability in each block. Task space variability increased after pre-test. An inverse u shape showed participants explored less in the later phase of the training. (B) Null space variability in each block. The null space did not change much through the learning.

### *Null space variability*

### Training phase

There was a significant main effect by block ( $F(1, 32) = 5.66, p = .02$ ), null space variability was lower in B6 compared to B1. There was no significant main effect by group ( $F(1,32) = 0.88, p = 0.6$ ) and no significant group x block interaction ( $F(3, 32) = 1.44, p = .25$ ) (Figure 4.15B).

### Test phase

There was a significant main effect by block ( $F(1, 32) = 5.51, p = .03$ ), null space variability was lower in post-test compared to pre-test. There was no significant main effect by group ( $F(1,32) = 0.75, p = .53$ ) and no significant group x block interaction ( $F(3, 32) = 1.12, p = .44$ ) (Figure 4.15B).

### ***Summary of experiment 2***

In Experiment 2, we added three variations of adaptive shaping to aim for a greater change of movement pattern. First, we modified the factor of memory in the algorithm, we wanted to see how participants reacted when including more previous trials in moving average (long adaptive group). Second, we used linear regression to calculate the thresholds, we wanted to see how exploration was performed with a wider search range (momentum group). Third, we modified both factors of memory and momentum together to see if both effects add up (long momentum group). The main findings that (i) all groups showed increased trunk velocity in the post-test, and (ii) although there were no statistical differences between the groups, the long-adaptive group seemed to perform the best relative to the adaptive group.

### **General Discussion**

The overall goal of this study was to examine the use of reinforcement feedback to change movement patterns in a complex motor skill with multiple degrees of freedom. We found two main results across both experiments – (i) reinforcement feedback in multi-DOF tasks was

not as effective as previously reported in adaptation experiments with no redundancy, and (ii) an adaptive schedule outperformed open-loop schedules, with evidence that a more conservative strategy for setting the threshold performed the best.

### ***Reinforcement in multi-DOF tasks***

Our first question addressed the issue of whether reinforcement learning would be successful in altering movement patterns in multi-DOF tasks. We found that reinforcement learning (across all the groups used here with different shaping schedules) was only partly successful with punishment rates ranging from 30-80%. In prior studies with motor adaptation tasks, reinforcement learning show reward rates of 100% (i.e. equivalent to a punishment rate of 0% in our case), indicating that participants were able to clearly use reinforcement feedback to modify their movement pattern to the desired level. However, as mentioned before, a critical limitation of these tasks is that they mostly used single DOF reaching tasks where the only exploration was along the angle of the reach.

The key challenge of reinforcement learning in multi-DOF tasks is exploration. When the exploration space is extremely small (as in single DOF tasks) reinforcement can easily guide the learner toward the solution, although there is some evidence that even in these tasks, there is a small proportion of individuals who fail to adapt (Chen et al., 2018). However, in the current task, with just 1 extra DOF, the two-dimensional DOF space was large enough that exploration was extremely difficult for participants. Moreover, even though the task was two dimensional, the exploration space could have been even higher dimensional. Experimental observations showed many participants exploring other body segments (such as the head or the other hand), indicating that participants were exploring a much higher dimensional space than what was specified in the task. Under these conditions, reinforcement learning, even with adaptive shaping

may not be sufficient to shift participants toward the desired movement pattern. Combining reinforcement feedback with other methods (such as demonstrations or attentional cues) may be necessary to improve exploration in high dimensional spaces.

### ***Shaping schedules***

To increase the efficiency of exploration in high dimensional spaces, shaping has been used in reinforcement learning (Ferster & Skinner, 1957). Shaping refers to providing additional rewards for making progress toward the desired solution and has been shown to be effective in animal experiments and also studies in artificial intelligence (Knox & Stone, 2009). Here we examined different shaping schedules in the context of altering movement patterns.

### ***Abrupt vs. Gradual***

In Experiment 1, we first compared two open-loop schedules - abrupt and gradual. The abrupt schedule, where the threshold is instantaneously raised to the desired level, is equivalent to no shaping (since there is no change in feedback until the desired movement pattern is reached), whereas the gradual group uses shaping (since small incremental changes in trunk velocity receive changes in feedback). Interestingly, although these two methods have had different effects on learning in a number of adaptation studies (Kagerer et al., 1997; Ludolph et al., 2017; Milner et al., 2018), we found that the abrupt and gradual group were not distinguishable in terms of the change of trunk velocity. However, the punishment rate suggested that they had two different routes to learning – the abrupt group, as expected, had high punishment rates throughout practice, indicating that exploration was not successful in determining the correct solution. However, the gradual group started off with low punishment rates but this rate kept increasing as the threshold increased, indicating a lack of ability to adapt to the feedback. One potential reason is that because the threshold in the gradual group increases

in an open-loop manner, once participants initially failed to adapt to the changing threshold, subsequent changes in threshold were even further away making it almost similar to the abrupt condition. These results suggest that open-loop shaping schedules are suboptimal in multi-DOF tasks because when exploration is slow and inefficient, even small changes in the threshold over trials can quickly become difficult to overcome.

### *Adaptive schedules*

To overcome this limitation of open-loop schedules, we also tested an adaptive group, where the threshold was based on the participants' performance. The punishment rate in this group was maintained around 50% throughout practice and was associated with the greatest increase in trunk velocity. These results support prior studies using adaptive schedules in motor adaptation (Therrien et al., 2018; Verstynen & Sabes, 2011) and suggest that the adaptive schedule created a condition where the exploration was at least moderately successful even in a multi-DOF task.

In Experiment 2, we further examined the by manipulating two other factors – memory and momentum. Both factors essentially varied the aggressiveness with which the threshold changed over practice. Increasing memory to include more past information made the threshold estimate more conservative, whereas increasing the momentum to examine the trend across the past information made the estimate more aggressive. From the results, we found that the best learning results were found for the conservative threshold estimate, as seen by the lower punishment rate and the higher trunk velocity in the post-test. These results suggest that a more conservative estimate gave participants some time to explore and settle into a new movement pattern.

The success of the more conservative method is likely tied to that in our task, participants had to move away from their ‘preferred’ coordination pattern which involved mostly hand motion. Again, unlike prior adaptation studies where the only change required is the angle of reach (which does not have strong preferred directions), the task in the current experiment resembles real-life contexts (such as changing a golf swing or rehabilitation of a movement pattern) where there is a strong preference for an existing coordination pattern. These highlight the importance of considering the dynamical systems view of coordination where coordination patterns do not exist on a ‘blank slate’, but have different stability properties (Schoner & Kelso, 1988; Sternad, 1998). Overall, our results suggest that in such cases, reinforcement feedback with conservative adaptive are most likely to result in better learning outcomes.

In summary, we found that reinforcement feedback can be used to change the movement patterns in a multi-DOF task. Adaptive schedules that modified reinforcement feedback based on participant performance had the greatest chance of modifying coordination patterns, especially when they were slow to create change. These results highlight the potential of reinforcement feedback in multi-DOF tasks and suggest that future work on adaptive schedules is needed to accelerate learning while still allowing participants to explore the space of possible solution.

## CHAPTER 5 GENERAL DISCUSSION

### Overall scope

The overall aim of this dissertation is to investigate how to guide participants to use alternative movement patterns to perform the same task by using reinforcement learning. Motor learning studies have mainly focused on the change of the task performance, with little focus on the change of movement patterns. However, studying the change of movement patterns is critical in contexts where different movement patterns can be used, but there may be advantages to using certain specific movement patterns. For example, in movement rehabilitation, the individuals with movement disabilities learn the correct way to perform the task to prevent further injuries. The process of learning different movement patterns is long and challenging because humans tend to use habitual movement patterns regardless of the pattern is optimal or not (De Rugy et al., 2012).

To approach this question, I used reinforcement feedback to guide participants to learn alternative movement patterns. Reinforcement learning is the theoretical framework that connects three experiments in this dissertation. In motor adaptation research, reinforcement learning was used as a form of reward and punishment learning. The learners either changed the movement patterns to pursue the rewards or prevent the punishments. Given that prior work on reinforcement has used tasks that only require small manipulations of well-learned movement patterns (e.g. changing the direction of a reach), an important contribution of the current work is to examine how these reinforcement paradigms generalize to tasks involving the coordination of multiple DOFs.

For all the three experiments, I used trunk-hand coordination as the multi-DOF task. Trunk-hand coordination is ubiquitous in our daily life, for example, when we reach, when we

throw, etc. The participants controlled the kinematics of the trunk and the hand to perform the task. With two degree of freedom in the system, there are many combinations of trunk and hand kinematics to solve the task. This setup provides the redundancy to study learning different movement patterns to perform the same task. Moreover, unlike bimanual movements where there are strong tendencies toward symmetry, using the trunk and the hand provided an opportunity to explore a larger range of coordination patterns.

### **Contributions of the dissertation**

The first contribution of this dissertation is to investigate the reinforcement learning protocol in multi-DOF tasks. Given that human movement requires coordination of large number of DOFs (such as joints and muscles), I extended methods that have been used mainly for adaptation studies (where there is no change in the underlying coordination) to tasks requiring coordination of multiple DOFs. Increasing the number of DOFs is both theoretically relevant (as it now creates possibilities for multiple movement patterns to perform the task) and practically relevant (as it relates to common motor learning contexts like learning a golf swing or learning to reach after a stroke). I provided a series of experiments to approach the problem of learning multi-DOF tasks with the focus on changing movement patterns.

The second contribution is to apply the concept of reinforcement learning and shaping to learn alternative movement patterns. The concept of shaping (or adapting reinforcement feedback based on individual performance) is widely used in many contexts of motor learning. For example, a coach adjusts the difficulty of the task based on the performance level of the athletes. Despite a general acceptance that such adaptive learning is a good strategy, how exactly shaping improve learning is not clear. My studies provide a systematic perspective to show how changing the parameters in shaping algorithm affects learning outcomes.

## **Online feedback vs delayed feedback**

One issue that I investigated in this dissertation concerns the timing of the feedback and how it affects learning of new coordination patterns. In experiment one, the reinforcement feedback was provided online during reaching. The participants got punished instantaneously if the trunk movement was smaller than the threshold. The experimental design showed reinforcement feedback successfully changed the movement patterns. However, such online feedback is limited in two ways – (a) it can only be used in slow positioning tasks where participants can react to the feedback during the movement, and (ii) it creates a confound in understanding the role of feedback because different participants receive different amounts of feedback (depending on how long they take to do the movement and how many errors they make).

To better quantify the reinforcement feedback, in experiment two I used a discrete task that participants received one feedback after each trial. This allows this paradigm to be used in fast movements and also allows for controlling the total amount of feedback during training. With this design, I could calculate the rate of receiving feedback to show how reinforcement rate related to exploration. Providing offline feedback solved the problem of comparing across participants but increased the difficulty of the task. The feedback was given after the whole trial so the participants did not have much opportunity to map the feedback to the movement pattern. Especially the participants did not know the feedback was generated only by trunk velocity so they needed more trials to understand the feedback.

Moreover, I used the reinforcement rate to discuss the exploration-exploitation tradeoff. Exploration- exploitation tradeoff is a fundamental problem to solve during learning. Providing online feedback worked better with the abrupt change of the threshold. The participants explored

extensively until they found the good solution. After this point, they exploit what they had learned to keep the good performance. On the contrary, providing offline feedback worked better with the change in the threshold was adaptive to the performance level of the learner. The exploration-exploitation ratio maintained at a certain level through the learning. The results showed that it is important to provide the task at a certain difficulty to have a good balance of exploration- exploitation ratio in learning.

In sum, providing online feedback was a possible protocol to change movement patterns but it had the limitation to show how the amount of feedback guided the exploration. Providing offline feedback provided a fair comparison between participants so how exploration guided learning was clearer.

### **Shaping reward/punishment during reinforcement**

Shaping is a common method to make sure the participants do not lose motivation and keep exploring, this term is from the early research of conditioning (Ferster & Skinner, 1957). The concept was also used in machine learning to help the agent to explore. There are two different ways to design shaping methods: 1) manipulating the task difficulty 2) manipulating the way to generate feedback. In the first type, the difficulty of the task changed during learning. For example, the algorithm lowers the difficulty of the task when the participants had trouble to receive the reward to help participants. In the visual-motor rotation tasks, participants receive the reward when they had the correct direction is an example of changing the actual task difficulty (Therrien et al., 2018).

The second type, participants received different feedback based on their behavior (Coltman et al., 2019). For example, providing graded reinforcement feedback so participants got different degrees of reward. The two types seem similar on the behavior level, but they

interpretation could be very different from how the central nervous system learns the task. The main difference between the two types is whether the brain learns the cost function of the task or not (Körding & Wolpert, 2004). If the task difficulty is changing, that means the brain needs to learn different cost functions trial by trial. This mechanism leads to model-free reinforcement learning (Haith & Krakauer, 2013). For the second type, the task difficulty remains the same, the brain can actually learn the cost function, and this is a form of model-based reinforcement learning. My experiments followed the second type where the actual task difficulty did not change whereas I manipulated the protocol to provide feedback. However, the theoretical framework explains the behavior well, there is still no empirical evidence to show the difference between these two types to provide shaping, especially in multi-DOF tasks. My experiment is a step toward answering this question.

These also raise many other interesting questions. There are many hyperparameters to play around to investigate how humans react to different shaping methods. For example, linear regression and moving average are both parametric methods to estimate the state of the participants. Non-parametric models like Gaussian Process (Nguyen-Tuong et al., 2009) can be used to fit the data. Changing the fitting model does not only have the chance to get different results in terms of the outcome but also provide a window into the learning mechanism.

### **Limitation and future direction**

One limitation of this dissertation is the lack of methods to show how the variability plays the role in reinforcement learning in multi-DOF tasks. Reinforcement learning is an open paradigm where the learners are encouraged to explore. More exploration often leads to higher variance both on the performance and movement pattern level. This type of variability is essential to learn new movement patterns. However, exploration causes outliers in the data and

conventional linear method is not robust when estimating variance with outliers. Running outlier detection is not always a possibility since these outliers can potentially have important information. This problem becomes more serious when the tasks have more DOFs that the learner needs to explore more to find the solution. Finding a good method that provides consistent estimation of the variance across different types of exploration pattern is critical to describe this type of learning. One proposed solution for future studies is to use probabilistic modeling to model the distribution of the whole landscape then estimate the expected value and variance. Now the “outliers” are weighted by the probability, i.e., higher probability means the learner followed a pattern. This gives us more flexibility to accommodate the sparsity of the data and returns a better estimation of the variability. After having a reliable estimation of variability, the next challenge is to separate motor noise and exploration from variability. In a single-DOF task, this could be done by assuming participants did not change their strategy if they did not receive reinforcement feedback, then all the variability was motor noise and the changes after receiving reinforcement feedback was exploration. With this assumption, a stochastic process such as particle filter (Therrien et al., 2016) was applied to separate exploration and noise. However, this assumption could be problematic in multi-DOF tasks due to exploration can happen on all the dimensions, it is difficult to quantify exploration when multiple dimensions are involved. The future direction of this line of research will be to develop some methods to capture the “exploration” in multi-DOF tasks so the link between reinforcement learning and exploration can be further shown in motor learning.

## **Conclusion**

Motor learning was mostly studied from the perspective of the change of performance but seldom focused on the change of movement patterns. Studying the change of movement patterns

provides additional information to describe how motor learning is implemented in the human nervous system. I used reinforcement learning as the theoretical framework to investigate how to guide participants using alternative movement patterns. In summary, this dissertation showed three main findings: 1) Reinforcement learning protocol could be applied to guide participants to use alternative movement patterns in a multi-DOF task. 2) Shaping outperformed non-shaping methods in shifting movement patterns in reinforcement learning protocol. 3) Changing parameters in shaping changed the performance of learning the alternative movement patterns.

## **REFERENCES**

## REFERENCES

- Berniker, M., & Kording, K. (2008). Estimating the sources of motor errors for adaptation and generalization. *Nature Neuroscience*, *11*(12), 1454–1461. <https://doi.org/10.1038/nn.2229>
- Bernstein, N. A. (1967). *The Co-ordination and regulation of movements*. Pergamon Press Ltd.
- Bhushan, N., & Shadmehr, R. (1999). Computational nature of human adaptive control during learning of reaching movements in force fields. *Biological Cybernetics*, *81*, 39–60.
- Chen, X., Holland, P., & Galea, J. M. (2018). The effects of reward and punishment on motor skill learning. *Current Opinion in Behavioral Sciences*, *20*, 83–88. <https://doi.org/10.1016/j.cobeha.2017.11.011>
- Cirstea, M. C., & Levin, M. F. (2000). Compensatory strategies for reaching in stroke. *Brain*, *123*(5), 940–953. <https://doi.org/10.1093/brain/123.5.940>
- Coltman, S. K., Cashaback, J. G. A., & Gribble, P. L. (2019). Both fast and slow learning processes contribute to savings following sensorimotor adaptation. *Journal of Neurophysiology*, *121*(4), 1575–1583. <https://doi.org/10.1152/jn.00794.2018>
- Criscimagna-Hemminger, S. E., Bastian, A. J., & Shadmehr, R. (2010). Size of Error Affects Cerebellar Contributions to Motor Learning. *Journal of Neurophysiology*, *103*(4), 2275–2284. <https://doi.org/10.1152/jn.00822.2009>
- Cusumano, J. P., & Cesari, P. (2006). Body-goal Variability Mapping in an Aiming Task. *Biological Cybernetics*, *94*(5), 367–379. <https://doi.org/10.1007/s00422-006-0052-1>
- De Maesschalck, R., Jouan-Rimbaud, D., & Massart, D. L. (2000). The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, *50*(1), 1–18. [https://doi.org/10.1016/S0169-7439\(99\)00047-7](https://doi.org/10.1016/S0169-7439(99)00047-7)
- De Rugy, A., Loeb, G. E., & Carroll, T. J. (2012). Muscle Coordination Is Habitual Rather than Optimal. *Journal of Neuroscience*, *32*(21), 7384–7391. <https://doi.org/10.1523/JNEUROSCI.5792-11.2012>
- Dhawale, A. K., Smith, M. A., & Ölveczky, B. P. (2017). The Role of Variability in Motor Learning. *Annual Review of Neuroscience*, *40*(1), 479–498. <https://doi.org/10.1146/annurev-neuro-072116-031548>
- Diedrichsen, J. (2007). Optimal Task-Dependent Changes of Bimanual Feedback Control and Adaptation. *Current Biology*, *17*(19), 1675–1679. <https://doi.org/10.1016/j.cub.2007.08.051>

- Diedrichsen, J., Shadmehr, R., & Ivry, R. B. (2010). The coordination of movement: Optimal feedback control and beyond. *Trends in Cognitive Sciences*, 14(1), 31–39. <https://doi.org/10.1016/j.tics.2009.11.004>
- Dingwell, J. B., John, J., & Cusumano, J. P. (2010). Do Humans Optimally Exploit Redundancy to Control Step Variability in Walking? *PLoS Computational Biology*, 6(7), e1000856. <https://doi.org/10.1371/journal.pcbi.1000856>
- Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. Appleton-Century-Crofts. <https://doi.org/10.1037/10627-000>
- Ficuciello, F., Falco, P., & Calinon, S. (2018). A Brief Survey on the Role of Dimensionality Reduction in Manipulation Learning and Control. *IEEE Robotics and Automation Letters*, 3(3), 2608–2615. <https://doi.org/10.1109/LRA.2018.2818933>
- Franklin, D. W., & Wolpert, D. M. (2011). Computational Mechanisms of Sensorimotor Control. *Neuron*, 72(3), 425–442. <https://doi.org/10.1016/j.neuron.2011.10.006>
- Galea, J. M., Mallia, E., Rothwell, J., & Diedrichsen, J. (2015). The dissociable effects of punishment and reward on motor learning. *Nature Neuroscience*, 18(4), 597–602. <https://doi.org/10.1038/nn.3956>
- Ganesh, G., Haruno, M., Kawato, M., & Burdet, E. (2010). Motor Memory and Local Minimization of Error and Effort, Not Global Optimization, Determine Motor Behavior. *Journal of Neurophysiology*, 104(1), 382–390. <https://doi.org/10.1152/jn.01058.2009>
- Gläscher, J., Daw, N., Dayan, P., & 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. <https://doi.org/10.1016/j.neuron.2010.04.016>
- Gløersen, Ø., Myklebust, H., Hallén, J., & Federolf, P. (2018). Technique analysis in elite athletes using principal component analysis. *Journal of Sports Sciences*, 36(2), 229–237. <https://doi.org/10.1080/02640414.2017.1298826>
- Guigon, E., Baraduc, P., & Desmurget, M. (2007). Computational Motor Control: Redundancy and Invariance. *Journal of Neurophysiology*, 97(1), 331–347. <https://doi.org/10.1152/jn.00290.2006>
- Haith, A. M., & Krakauer, J. W. (2013). Model-Based and Model-Free Mechanisms of Human Motor Learning. In M. J. Richardson, M. A. Riley, & K. Shockley (Eds.), *Progress in Motor Control* (Vol. 782, pp. 1–21). Springer New York. [https://doi.org/10.1007/978-1-4614-5465-6\\_1](https://doi.org/10.1007/978-1-4614-5465-6_1)
- Herzfeld, D. J., & Shadmehr, R. (2014). Motor variability is not noise, but grist for the learning mill. *Nature Neuroscience*, 17(2), 149–150. <https://doi.org/10.1038/nn.3633>

- Huang, V. S., & Shadmehr, R. (2009). Persistence of Motor Memories Reflects Statistics of the Learning Event. *Journal of Neurophysiology*, 102(2), 931–940.  
<https://doi.org/10.1152/jn.00237.2009>
- Izawa, J., & Shadmehr, R. (2011). Learning from Sensory and Reward Prediction Errors during Motor Adaptation. *PLoS Computational Biology*, 7(3), e1002012.  
<https://doi.org/10.1371/journal.pcbi.1002012>
- Jenkins, O. C., & Matarić, M. J. (2004). A spatio-temporal extension to Isomap nonlinear dimension reduction. *Twenty-First International Conference on Machine Learning - ICML '04*, 56. <https://doi.org/10.1145/1015330.1015357>
- Jordan, M. I., & Rumelhart, D. E. (1992). Forward Models: Supervised Learning with a Distal Teacher. *Cognitive Science*, 16(3), 307–354.  
[https://doi.org/10.1207/s15516709cog1603\\_1](https://doi.org/10.1207/s15516709cog1603_1)
- Kagerer, F. A., Contreras-Vidal, J. L., & Stelmach, G. E. (1997). Adaptation to gradual as compared with sudden visuo-motor distortions. *Experimental Brain Research*, 115(3), 557–561.
- Klassen, J., Tong, C., & Flanagan, J. R. (2005). Learning and recall of incremental kinematic and dynamic sensorimotor transformations. *Experimental Brain Research*, 164(2), 250–259.  
<https://doi.org/10.1007/s00221-005-2247-4>
- Knox, W. B., & Stone, P. (2009). Interactively shaping agents via human reinforcement: The TAMER framework. *Proceedings of the Fifth International Conference on Knowledge Capture - K-CAP '09*, 9. <https://doi.org/10.1145/1597735.1597738>
- Konczak, J., vander Velden, H., & Jaeger, L. (2009). Learning to Play the Violin: Motor Control by Freezing, Not Freeing Degrees of Freedom. *Journal of Motor Behavior*, 41(3), 243–252. <https://doi.org/10.3200/JMBR.41.3.243-252>
- Körding, K. P., & Wolpert, D. M. (2004). The loss function of sensorimotor learning. *Proceedings of the National Academy of Sciences of the United States of America*, 101(26), 9839–9842.
- Krakauer, J. W., & Mazzoni, P. (2011). Human sensorimotor learning: Adaptation, skill, and beyond. *Current Opinion in Neurobiology*, 21(4), 636–644.  
<https://doi.org/10.1016/j.conb.2011.06.012>
- Krakauer, J. W., Pine, Z. M., Ghilardi, M.-F., & Ghez, C. (2000). Learning of Visuomotor Transformations for Vectorial Planning of Reaching Trajectories. *Journal of Neuroscience*, 20(23), 8916–8924.

- Latash, M. L. (2012). The bliss (not the problem) of motor abundance (not redundancy). *Experimental Brain Research*, 217(1), 1–5. <https://doi.org/10.1007/s00221-012-3000-4>
- Latash, M. L., Scholz, J. P., & Schöner, G. (2002). Motor control strategies revealed in the structure of motor variability. *Exercise and Sport Sciences Reviews*, 30(1), 26–31.
- Levin, M. F., Kleim, J. A., & Wolf, S. L. (2009). What Do Motor “Recovery” and “Compensation” Mean in Patients Following Stroke? *Neurorehabilitation and Neural Repair*, 23(4), 313–319. <https://doi.org/10.1177/1545968308328727>
- Ludolph, N., Giese, M. A., & Ilg, W. (2017). Interacting Learning Processes during Skill Acquisition: Learning to control with gradually changing system dynamics. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-13510-0>
- Martin, J. R., Zatsiorsky, V. M., & Latash, M. L. (2011). Multi-finger interaction during involuntary and voluntary single finger force changes. *Experimental Brain Research*, 208(3), 423–435. <https://doi.org/10.1007/s00221-010-2492-z>
- Martin, T. A., Keating, J. G., Goodkin, H. P., Bastian, A. J., & Thach, W. T. (1996). Throwing while looking through prisms: I. Focal olivocerebellar lesions impair adaptation. *Brain*, 119(4), 1183–1198. <https://doi.org/10.1093/brain/119.4.1183>
- Mehler, D. M. A., Reichenbach, A., Klein, J., & Diedrichsen, J. (2017). Minimizing endpoint variability through reinforcement learning during reaching movements involving shoulder, elbow and wrist. *PLOS ONE*, 12(7), e0180803. <https://doi.org/10.1371/journal.pone.0180803>
- Michaelson, S. M., Dannenbaum, R., & Levin, M. F. (2006). Task-Specific Training With Trunk Restraint on Arm Recovery in Stroke: Randomized Control Trial. *Stroke*, 37(1), 186–192. <https://doi.org/10.1161/01.STR.0000196940.20446.c9>
- Milner, T. E., Firouzi Mehr, Z., Babadi, S., & Ostry, D. J. (2018). Different adaptation rates to abrupt and gradual changes in environmental dynamics. *Experimental Brain Research*, 236(11), 2923–2933. <https://doi.org/10.1007/s00221-018-5348-6>
- Müller, H., & Sternad, D. (2004). Decomposition of Variability in the Execution of Goal-Oriented Tasks: Three Components of Skill Improvement. *Journal of Experimental Psychology: Human Perception and Performance*, 30(1), 212–233. <https://doi.org/10.1037/0096-1523.30.1.212>
- Murillo, D. B., Sánchez, C. C., Moreside, J., Vera-García, F. J., & Moreno, F. J. (2017). Can the structure of motor variability predict learning rate? *Journal of Experimental Psychology: Human Perception and Performance*, 43(3), 596–607. <https://doi.org/10.1037/xhp0000303>

- Mussa-Ivaldi, F. A., Casadio, M., Danziger, Z. C., Mosier, K. M., & Scheidt, R. A. (2011). Sensory motor remapping of space in human-machine interfaces. In *Progress in Brain Research* (Vol. 191, pp. 45–64). Elsevier. <https://doi.org/10.1016/B978-0-444-53752-2.00014-X>
- Neilson, P. D. (1993). The problem of redundancy in movement control: The adaptive model theory approach. *Psychological Research*, 55(2), 99–106. <https://doi.org/10.1007/BF00419640>
- Newell, K. M. (1986). Constraints on the development of coordination. *Motor Development in Children: Aspects of Coordination and Control*.
- Newell, K. M., Broderick, M. P., Deutsch, K. M., & Slifkin, A. B. (2003). Task goals and change in dynamical degrees of freedom with motor learning. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2), 379–387. <https://doi.org/10.1037/0096-1523.29.2.379>
- Newell, K. M., & Corcos, D. M. (1993). *Variability and Motor Control*. Human Kinetics Publishers.
- Newell, K. M., & Vaillancourt, D. E. (2001). Dimensional change in motor learning. *Human Movement Science*, 20(4–5), 695–715. [https://doi.org/10.1016/S0167-9457\(01\)00073-2](https://doi.org/10.1016/S0167-9457(01)00073-2)
- Nguyen-Tuong, D., Seeger, M., & Peters, J. (2009). Model Learning with Local Gaussian Process Regression. *Advanced Robotics*, 23(15), 2015–2034. <https://doi.org/10.1163/016918609X12529286896877>
- Nikooyan, A. A., & Ahmed, A. A. (2015). Reward feedback accelerates motor learning. *Journal of Neurophysiology*, 113(2), 633–646. <https://doi.org/10.1152/jn.00032.2014>
- Ranganathan, R., Wieser, J., Mosier, K. M., Mussa-Ivaldi, F. A., & Scheidt, R. A. (2014). Learning Redundant Motor Tasks with and without Overlapping Dimensions: Facilitation and Interference Effects. *Journal of Neuroscience*, 34(24), 8289–8299. <https://doi.org/10.1523/JNEUROSCI.4455-13.2014>
- Ranganathan, R., Wang, R., Gebara, R., & Biswas, S. (2017). Detecting Compensatory Trunk Movements in Stroke Survivors using a Wearable System. *Proceedings of the 2017 Workshop on Wearable Systems and Applications - WearSys '17*, 29–32. <https://doi.org/10.1145/3089351.3089353>
- Reynolds, G. S. (1961). Relativity of response rate and reinforcement frequency in a multiple schedule. *Journal of the Experimental Analysis of Behavior*, 4(2), 179–184.

- Rosenbaum, D. A., & Jorgensen, M. J. (1992). Planning macroscopic aspects of manual control. *Human Movement Science*, 11(1–2), 61–69. [https://doi.org/10.1016/0167-9457\(92\)90050-L](https://doi.org/10.1016/0167-9457(92)90050-L)
- Safonova, A., Hodgins, J. K., & Pollard, N. S. (2004). Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces. *ACM Transactions on Graphics (ToG)*, 23(3), 514–521.
- Schaal, S., Ijspeert, A., & Billard, A. (2003). Computational approaches to motor learning by imitation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358(1431), 537–547. <https://doi.org/10.1098/rstb.2002.1258>
- Schaal, S., & Schweighofer, N. (2005). Computational motor control in humans and robots. *Current Opinion in Neurobiology*, 15(6), 675–682. <https://doi.org/10.1016/j.conb.2005.10.009>
- Scholz, J. P., & Schöner, G. (1999). The uncontrolled manifold concept: Identifying control variables for a functional task. *Experimental Brain Research*, 126(3), 289–306. <https://doi.org/10.1007/s002210050738>
- Scholz, J.P., Danion, F., Latash, M. L., & Schöner, G. (2002). Understanding finger coordination through analysis of the structure of force variability. *Biological Cybernetics*, 86(1), 29–39.
- Schöner, G., & Kelso, J. (1988). Dynamic pattern generation in behavioral and neural systems. *Science*, 239(4847), 1513–1520. <https://doi.org/10.1126/science.3281253>
- Selinger, J. C., O'Connor, S. M., Wong, J. D., & Donelan, J. M. (2015). Humans Can Continuously Optimize Energetic Cost during Walking. *Current Biology*, 25(18), 2452–2456. <https://doi.org/10.1016/j.cub.2015.08.016>
- Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error Correction, Sensory Prediction, and Adaptation in Motor Control. *Annual Review of Neuroscience*, 33(1), 89–108. <https://doi.org/10.1146/annurev-neuro-060909-153135>
- Singh, P., Jana, S., Ghosal, A., & Murthy, A. (2016). Exploration of joint redundancy but not task space variability facilitates supervised motor learning. *Proceedings of the National Academy of Sciences*, 113(50), 14414–14419. <https://doi.org/10.1073/pnas.1613383113>
- Skinner, B. F. (1938). *The Behavior of Organisms*. Appleton-Century-Crofts. New York.
- Stergiou, N., & Decker, L. M. (2011). Human movement variability, nonlinear dynamics, and pathology: Is there a connection? *Human Movement Science*, 30(5), 869–888. <https://doi.org/10.1016/j.humov.2011.06.002>

- Sternad, D. (1998). Hot Topics in Motor Control and Learning: A Dynamic Systems Perspective to Perception and Action. *Research Quarterly for Exercise and Sport*, 69(4), 319–325. <https://doi.org/10.1080/02701367.1998.10607705>
- Sutton, R. S., & Barto, A. G. (2017). *Reinforcement Learning: An Introduction* (2nd ed.). The MIT Press.
- Therrien, A. S., Wolpert, D. M., & Bastian, A. J. (2016). Effective reinforcement learning following cerebellar damage requires a balance between exploration and motor noise. *Brain*, 139(1), 101–114. <https://doi.org/10.1093/brain/awv329>
- Therrien, A. S., Wolpert, D. M., & Bastian, A. J. (2018). Increasing Motor Noise Impairs Reinforcement Learning in Healthy Individuals. *Eneuro*, 5(3), ENEURO.0050-18.2018. <https://doi.org/10.1523/ENEURO.0050-18.2018>
- Thorp, E. B., Kording, K. P., & Mussa-Ivaldi, F. A. (2017). Using noise to shape motor learning. *Journal of Neurophysiology*, 117(2), 728–737. <https://doi.org/10.1152/jn.00493.2016>
- Todorov, E., & Ghahramani, Z. (2003). Unsupervised learning of sensory-motor primitives. *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)*, 1750–1753. <https://doi.org/10.1109/IEMBS.2003.1279744>
- Todorov, E., & Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, 5(11), 1226–1235. <https://doi.org/10.1038/nn963>
- Tseng, Y., Diedrichsen, J., Krakauer, J. W., Shadmehr, R., & Bastian, A. J. (2007). Sensory Prediction Errors Drive Cerebellum-Dependent Adaptation of Reaching. *Journal of Neurophysiology*, 98(1), 54–62. <https://doi.org/10.1152/jn.00266.2007>
- Vereijken, B., Emmerik, R. E. A. van, Whiting, H. T. A., & Newell, K. M. (1992). Free(z)ing Degrees of Freedom in Skill Acquisition. *Journal of Motor Behavior*, 24(1), 133–142. <https://doi.org/10.1080/00222895.1992.9941608>
- Verstynen, T., & Sabes, P. N. (2011). How Each Movement Changes the Next: An Experimental and Theoretical Study of Fast Adaptive Priors in Reaching. *Journal of Neuroscience*, 31(27), 10050–10059. <https://doi.org/10.1523/JNEUROSCI.6525-10.2011>
- Wang, L., & Suter, D. (2007). Learning and Matching of Dynamic Shape Manifolds for Human Action Recognition. *IEEE Transactions on Image Processing*, 16(6), 1646–1661. <https://doi.org/10.1109/TIP.2007.896661>
- Witte, K., Ganter, N., Baumgart, C., & Peham, C. (2010). Applying a principal component analysis to movement coordination in sport. *Mathematical and Computer Modelling of Dynamical Systems*, 16(5), 477–488. <https://doi.org/10.1080/13873954.2010.507079>

- Wolpert, D. M., Diedrichsen, J., & Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience*, 12(12), 739. <https://doi.org/10.1038/nrn3112>
- Wolpert, D. M., & Flanagan, J. R. (2016). Computations underlying sensorimotor learning. *Current Opinion in Neurobiology*, 37, 7–11. <https://doi.org/10.1016/j.conb.2015.12.003>
- Wolpert, D. M., Ghahramani, Z., & Flanagan, J. R. (2001). Perspectives and problems in motor learning. *Trends in Cognitive Sciences*, 5(11), 487–494. [https://doi.org/10.1016/S1364-6613\(00\)01773-3](https://doi.org/10.1016/S1364-6613(00)01773-3)
- Wolpert, D. M., Ghahramani, Z., & Jordan, M. (1995). An internal model for sensorimotor integration. *Science*, 269(5232), 1880–1882. <https://doi.org/10.1126/science.7569931>
- Wu, H. G., Miyamoto, Y. R., Castro, L. N. G., Ölveczky, B. P., & Smith, M. A. (2014). Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nature Neuroscience*, 17(2), 312–321. <https://doi.org/10.1038/nn.3616>
- Wulf, G., & Weigelt, C. (1997). Instructions about Physical Principles in Learning a Complex Motor Skill: To Tell or Not to Tell.... *Research Quarterly for Exercise and Sport*, 68(4), 362–367. <https://doi.org/10.1080/02701367.1997.10608018>
- Yang, J., & Scholz, J. P. (2005). Learning a throwing task is associated with differential changes in the use of motor abundance. *Experimental Brain Research*, 163(2), 137–158. <https://doi.org/10.1007/s00221-004-2149-x>