INTERPLAY OF SENSORIMOTOR AND REWARD LEARNING IN HUMAN MOTOR CONTROL

Tjaša Kunavar

Doctoral Dissertation
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia

Supervisor: Prof. Dr. Jan Babič, IPS and Jožef Stefan Institute, Ljubljana, Slovenia

Evaluation Board:
Prof. Dr. Bojan Nemec, Chair, IPS and Jožef Stefan Institute, Ljubljana, Slovenia
Prof. Dr. Aleš Holobar, Member, University of Maribor, Maribor, Slovenia
Prof. Dr. Nejc Šarabon, Member, University of Primorska, Koper, Slovenia

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL



Tjaša Kunavar

INTERPLAY OF SENSORIMOTOR AND REWARD LEARNING IN HUMAN MOTOR CONTROL

Doctoral Dissertation

MEDSEBOJNO DELOVANJE SENZOMOTORIČNEGA UČENJA IN UČENJA Z NAGRADO PRI ČLOVEŠKI MOTORIČNI KONTROLI

Doktorska disertacija

Supervisor: Prof. Dr. Jan Babič

Ljubljana, Slovenia, September 2026

To my mom, who is forever in my heart.

Acknowledgments

I would like to express my gratitude to everyone who helped me during the time I spent working on this dissertation.

First, I thank my supervisor prof. dr. Jan Babič for his invaluable guidance and feedback given to me over the years. I sincerely thank him for being there when I needed a consultation or just friendly advice. I would like to thank my friends and colleagues that supported me and created a fun and inspiring working environment. I am grateful to my family and friends for always believing in me and supporting me throughout my studies.

The research in this thesis was funded by the Slovenian Research Agency (Research Core Funding No. P2-0076).

Abstract

Human motor learning is a multifaceted process shaped by distinct error signals, cognitive strategies, and ecological aspects. This thesis investigates the underlying mechanisms of motor adaptation through a series of studies examining how sensory prediction error and reward prediction error contribute to motor learning and control. Furthermore, it investigates how different error signals contribute to motor learning in environments where adaptation must balance performance optimization with risk mitigation.

Sensorimotor adaptation, driven by sensory prediction error, and reward-based learning, guided by reward prediction error, were explored using visuomotor and force-field perturbation paradigms. Findings reveal that while both learning modes enable adaptation to novel disturbances, only sensorimotor adaptation leads to robust internal model updates, which manifest as pronounced aftereffects indicative of implicit learning. In contrast, reward learning operates predominantly through explicit strategies, allowing rapid adaptation to unpredictable perturbations without inducing internal model changes.

Beyond traditional motor learning paradigms, this thesis introduces an ecologically grounded framework for motor control, emphasizing the adaptive importance of risk mitigation and enhancing efficiency in dynamic environments. Participants performing whole-body movements under novel force perturbations exhibited a high-level ecological control strategy that alternates between safety and efficiency depending on whether the task was successful or not. This dynamic switching, consistent with a win-stay/lose-shift policy, is governed by an integration of internal model adaptation, policy learning, and adaptive feedback control within a hierarchical reinforcement learning framework.

Collectively, these findings offer a unified account of human motor learning, emphasizing the complementary roles of sensorimotor and reward learning, and the influence of ecological factors in shaping adaptive motor behaviour in real-world contexts.

Povzetek

Človeško motorično učenje je večplastni proces, ki ga oblikujejo različni signali napak, kognitivne strategije in ekološki vidiki. Ta doktorska disertacija raziskuje osnovne mehanizme motorične prilagoditve skozi niz študij, ki preučujejo, kako napaka senzorične napovedi in napaka napovedi nagrade prispevata k učenju in nadzoru gibanja. Poleg tega preučuje, kako različni signali napake prispevajo k motoričnemu učenju v okoljih, kjer je potrebno uravnotežiti optimizacijo učinkovitosti in zmanjševanje tveganja.

Senzorimotorično učenje, ki temelji na napaki senzorične napovedi, in motorično učenje z nagrado, ki temelji na napaki napovedi nagrade, sta bila raziskana z uporabo paradigem vidnih rotacij in perturbacij s silo. Ugotovitve kažejo, da čeprav oba načina učenja omogočata prilagoditev na nove motnje, le senzorimotorična prilagoditev povzroči spremembe notranjega modela, kar se odraža v učinkih, ki vztrajajo po odstranitvi motnje in so značilni za implicitno učenje. Nasprotno pa se učenje z nagrado večinoma opira na eksplicitne strategije, ki omogočajo hitro prilagoditev na nepredvidljive perturbacije brez sprememb notranjega modela.

Disertacija uvaja ekološko utemeljen okvir motoričnega nadzora, ki presega omejitve tradicionalnih paradigem motoričnega učenja ter poudarja prilagoditveni pomen zmanjševanja tveganja in izboljševanja učinkovitosti v dinamičnih okoljih. Udeleženci, ki so izvajali gibe celega telesa pod vplivom novih perturbacij, so izkazali visokonivojsko ekološko strategijo nadzora, ki glede na uspešnost izvedbe naloge izmenjuje med varnostjo in učinkovitostjo. To dinamično preklapljanje, skladno z načelom ostani ob uspehu/spremeni ob neuspehu (angl. win-stay/lose-shift), poteka prek integracije sprememb notranjega modela, učenja strategije in prilagodljivega regulacijskega sistema v okviru hierarhičnega spodbujevanega učenja.

Skupno ti izsledki ponujajo enoten vpogled na človeško motorično učenje, pri čemer poudarjajo dopolnjujočo vlogo implicitnih in eksplicitnih procesov ter vpliv ekoloških dejavnikov pri oblikovanju prilagodljivega motoričnega vedenja v realnih kontekstih.

Contents

List of Figures **xv**

List of Tables **xvii**

Abbreviations **xix**

1 Introduction **1**
1.1 Introduction and Description of the Problem 1
1.1.1 Introduction to Human Motor Control and Learning 1
1.1.2 Feedforward and Feedback Motor Control 1
1.1.3 Human Motor Learning 2
1.1.3.1 Use-Dependent Learning 3
1.1.3.2 Sensorimotor Learning 3
1.1.3.3 Reward Learning 4
1.1.4 Experimental Paradigms 4
1.1.4.1 Adaptation Process 5
1.2 Thesis Purpose, Goals and Hypothesis 6
1.2.1 Purpose of the Dissertation 6
1.2.2 Goals of the Dissertation 6
1.2.3 Hypothesis 6
1.3 Thesis Structure 7

2 Explicit Learning Based on Reward Prediction Error Facilitates Agile Motor Adaptations **9**
2.1 Introduction 9
2.2 Results 10
2.2.1 Part 1: Consistent Direction of Perturbation 12
2.2.2 Part 2: Alternating Direction of Perturbation 14
2.3 Discussion 16
2.4 Methods 18
2.4.1 Participants 18
2.4.2 Experiment 18
2.4.3 Experimental Protocol 19
2.4.3.1 Consistent Direction of Perturbation 19
2.4.3.2 Alternating Direction of Perturbation 19
2.4.4 Data Analysis 20

3 Distinct Error Signals and Their Impact on Motor Adaptation: An Examination Through Force-Field Perturbation and Contextual Switching **21**
3.1 Introduction 21
3.2 Results 23

3.2.1 Part 1: Consistent Direction of Perturbation 26
3.2.2 Part 2: Alternating Direction of Perturbation 30
3.3 Discussion 33
3.4 Methods 36
3.4.1 Participants 36
3.4.2 Experiment 37
3.4.3 Experimental Protocol 38
3.4.3.1 Consistent Direction of Perturbation 38
3.4.3.2 Alternating Direction of Perturbation 38
3.4.4 Data Analysis 39

4 Success-Efficient /Failure-Safe Strategy for Hierarchical Reinforcement Motor Learning 41
4.1 Introduction 41
4.2 Results 43
4.2.1 Ecological Control: Success-Efficient/Failure-Safe Learning Strategy 45
4.2.2 Motor Plan Learning 47
4.2.3 Internal Model Learning 49
4.2.4 Feedback Gains are Manipulated Based on Success and Failure 50
4.3 Discussion 51
4.3.1 Balancing Safety and Efficiency 51
4.3.2 Modularity and Hierarchy in Motor Learning 53
4.3.3 Interplay of Internal Model Adaptation and Feedback Contributions 54
4.3.4 Limitations 55
4.4 Conclusion 55
4.5 Methods 56
4.5.1 Participants 56
4.5.2 Experimental Procedure 56
4.5.3 Measures and Statistical Analysis 57

5 Computational Model Integrating Sensorimotor and Reward Learning 61
5.1 Introduction 61
5.2 Iterative Learning Control Computational Model 61
5.2.1 ILC Model for Visuomotor Rotation 62
5.2.1.1 Simulation Setup 65
5.2.1.2 Simulation Results 66
5.2.2 Adaptive ILC Model for Visuomotor Rotation 70
5.2.2.1 Simulation Setup 72
5.2.2.2 Simulation Results 74
5.2.3 Adaptive ILC Model for Force-Field Perturbation 75
5.2.3.1 Simulation Setup 76
5.2.3.2 Simulation Results 77
5.3 Discussion 79

6 Conclusions 81

References 83

Bibliography 95

Biography 97

List of Figures

Figure 1.1: Feedforward and feedback control. 2
Figure 2.1: Experiment (visuomotor rotation). 11
Figure 2.2: Movements during consistent direction of perturbation. 13
Figure 2.3: Movements during random direction of perturbation. 15
Figure 3.1: Experiment (force field perturbation). 23
Figure 3.2: Kinematic error throughout the experiment with a single target location. 25
Figure 3.3: Kinematic error throughout the experiment with three target location. 25
Figure 3.4: Adaptation to consistent force-field perturbation. 27
Figure 3.5: Force compensation during consistent direction of perturbation. 29
Figure 3.6: After-effects following adaptation to consistent direction of perturbation. 30
Figure 3.7: Adaptation to random force-field perturbation. 31
Figure 3.8: Force compensation during random direction of perturbation. 32
Figure 4.1: Experimental setup and centre of mass adaptation dynamics. 44
Figure 4.2: Representative centre of mass trajectories and adaptation metrics. 46
Figure 4.3: Adaptation dynamics during perturbation and de-adaptation. 47
Figure 4.4: Progression of Smoothness and Co-Contraction. 49
Figure 5.1: Architecture of the computational model. 62
Figure 5.2: Simulation of different learning processes. 67
Figure 5.3: Simulation of a combination of sensorimotor and explicit reward learning. 70
Figure 5.4: Architecture of the adaptive computational model. 71
Figure 5.5: Simulation results of adaptive ILC model for visuomotor rotation. 74
Figure 5.6: Simulation results of adaptive ILC model for force-field perturbation. 78

List of Tables

Table 4.1: Statistical comparison of adaptation plateaus. 50

Abbreviations

ANOVA ... Analysis of Variance
CCW ... Counter-Clockwise
COM ... Centre of Mass
CW ... Clockwise
EMG ... Electromyography
EPE ... End-point error group
ERR ... Full visual error group
ILC ... Iterative Learning Control
RMS ... Root Mean Square
RWD ... Binary reward group
SD ... Standard Deviation
SEM ... Standard Error of Mean

Chapter 1

Introduction

1.1 Introduction and Description of the Problem

1.1.1 Introduction to Human Motor Control and Learning

Motor control refers to the processes by which humans plan, initiate, and coordinate voluntary movements, integrating sensory information with motor commands to produce purposeful actions. Humans are capable of highly sophisticated movements, effortlessly adapting to new environments and tasks (Scott, 2004). To produce desired movements, nervous system must learn how to move and have the ability to modify motor commands appropriately (Spampinato & Celnik, 2021). Generating such complex movements is a demanding process that requires precise control of the musculoskeletal system, guided by continuous sensory feedback(Haith & Krakauer, 2013b). To initiate a movement, the brain generates motor commands that are translated into muscle activations, which are then converted into purposeful movements through the peripheral motor system (Scott, 2004). Due to nonlinear dynamics of our bodies, each movement requires a complex sequence of motor commands to achieve the desired outcome (Haith & Krakauer, 2013b). There is no unique set of controls to achieve a specific goal. This leads to a complex motor system, but also allows for movements to be achieved more flexibly and robustly (Haith & Krakauer, 2013b).

1.1.2 Feedforward and Feedback Motor Control

Human motor control integrates feedforward mechanisms, which predict and plan motor actions in advance to achieve coordinated movement, and feedback mechanisms, which correct movements based on sensory feedback (Jordan, 1996).

Feedforward or predictive control consists of motor commands that are determined in advance based on our internal models and desired movement (Kawato, 1999). Internal models are neural representations of the relationship between the motor commands and the movement (D. W. Franklin et al., 2008). These models, believed to reside partly in the cerebellum, mimic the input/output characteristics—or their inverses—of the motor system. They predict the sensory consequences of movement and generate feedforward motor commands, enabling precise and efficient motor control (D. W. Franklin et al., 2008; Kawato, 1999; Kawato et al., 1987; Miall & Wolpert, 1996; Wolpert et al., 1998).

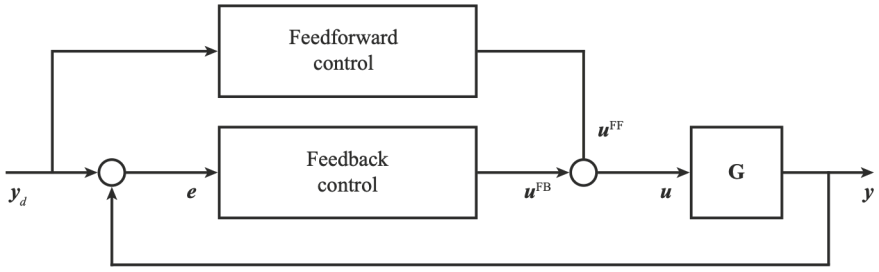


Figure 1.1: Feedforward and feedback control. Block diagram showing feedforward and feedback motor control, where control signal $\boldsymbol{u}$ is a sum of feedforward controller $\boldsymbol{u}^{\text{FF}}$, and feedback controller $\boldsymbol{u}^{\text{FB}}$. It directly generates the motor commands to the plant G. Feedback control is updated every time step based on the error $\boldsymbol{e}$.

Feedforward or predictive control consists of motor commands that are determined in advance based on our internal models and desired movement (Kawato, 1999). Internal models are neural representations of the relationship between the motor commands and the movement (D. W. Franklin et al., 2008). These models, believed to reside partly in the cerebellum, mimic the input/output characteristics—or their inverses—of the motor system. They predict the sensory consequences of movement and generate feedforward motor commands, enabling precise and efficient motor control (D. W. Franklin et al., 2008; Kawato, 1999; Kawato et al., 1987; Miall & Wolpert, 1996; Wolpert et al., 1998).

Another way to control our movement is through feedback or reactive control, where we can adapt the ongoing motor commands based on the sensorimotor information and the desired movement (D. W. Franklin & Wolpert, 2011; Kawato, 1999). Forward internal models can predict sensory consequences from efference copies of issued motor commands (Kawato, 1999). If there is a discrepancy between the efference copy and the sensorimotor information, motor commands need to account for these differences. Feedback control allows for a quick and responsive adaptation in our movement.

1.1.3 Human Motor Learning

The ability to learn the knowledge necessary to perform a variety of motor behaviours and quickly adapt them to the changing dynamics of the environment is a key aspect of human behaviour (Diedrichsen et al., 2010; D. W. Franklin & Wolpert, 2011; Scott, 2004; Spampinato & Celnik, 2021). Motor adaptation is a specific type of behavioral adjustment where an already well-practiced movement is fine-tuned to maintain performance despite changes in the environment or the body (Krakauer et al., 2019). Adaptation can be viewed as a process where the nervous system learns to predict the effects of a novel environment and searches for a better movement to minimize cost and maximize reward in that environment (Izawa et al., 2008). The goal seems to be to maximize performance, rather than cancelling changes in the new environment and returning movements to previous conditions (Izawa et al., 2008).

There are distinct learning processes underlying motor learning, including sensorimotor learning, reward learning and use-dependent learning (D. W. Franklin & Wolpert, 2011;

Spampinato & Celnik, 2021). Each form of learning contributes to a specific motor learning process to varying degrees, and typically, multiple processes work together rather than occurring in isolation (Spampinato & Celnik, 2021). While use-dependent learning occurs through repetition in the absence of systematic errors, sensorimotor and reward learning occur in the presence of sensory or reward prediction errors. Studies have demonstrated that relying on either type of error can yield similar levels of motor adaptation (Izawa & Shadmehr, 2011; Nikooyan & Ahmed, 2015; Palidis et al., 2019; Therrien et al., 2016). However, different types of error have distinct impacts on generalization, de-adaptation, and memory retention (Abe et al., 2011; Izawa & Shadmehr, 2011; Kunavar et al., 2023; K. Lee et al., 2018; Therrien et al., 2016).

1.1.3.1 Use-Dependent Learning

Use-dependent learning is considered a basic and goal independent way of motor learning (Mawase et al., 2017) and refers to the neural or behavioural changes that are induced through the simple repetition of movements (Bütefisch et al., 2000; Classen et al., 1998). Performing consistent movements in the absence of systematic errors induces use-dependent plastic changes in the motor cortex (Diedrichsen et al., 2010; Mawase et al., 2017). This affects future movements to become more similar to the previous movement (Diedrichsen et al., 2010). Use-dependent learning can explain why the motor system arrives at seemingly nonoptimal solutions when adapting to novel dynamics (Diedrichsen et al., 2010).

1.1.3.2 Sensorimotor Learning

Sensorimotor learning is a process by which the nervous system refines motor actions through the integration of sensory feedback and motor commands. It is driven by the detection and correction of sensory prediction errors—discrepancies between expected and actual sensory outcomes of movements. Sensory prediction errors are based on the sensory consequence of movement, observed by our sensory organs (e.g., vision, proprioception). The outcome of the movement from the sensorimotor system is compared to the desired or predicted outcome and the error between them is used to adjust the consequent motor commands (D. W. Franklin & Wolpert, 2011). When the sensory feedback is available, adaptation is driven largely by sensory prediction error (Izawa & Shadmehr, 2011).

Sensorimotor learning or error-based learning allows for alterations to the internal model through a feedforward model which maps motor commands with corresponding sensory consequences (Kawato, 1999; Shadmehr et al., 2010; Taylor et al., 2014) or through policy learning (Hadjiosif et al., 2021). If relationship between motor commands and movement is no longer valid, the internal model is updated, so that new motor commands are able to produce the desired movement. Sensorimotor learning is thought to be an implicit process where alterations of the internal model occur without conscious awareness (Mazzoni & Krakauer, 2006a). It takes place over a longer period of time, cannot be reversed immediately and can result in large and prolonged aftereffects reflecting the changes to the internal model during previous adaptation (Hegele & Heuer, 2010; Heuer & Hegele, 2008; Izawa & Shadmehr, 2011; Miyamoto et al., 2020).

1.1.3.3 Reward Learning

Reward learning is a process where we learn motor behaviours based on the feedback whether a movement was successful or a failure. It is driven by the reward prediction error which is the difference between the goal and the observed outcome. A failure to achieve an internally determined task goal in a goal directed movement is defined as a task error (L.-A. Leow et al., 2020; Schultz, 2017). While some tasks might be assessed in binary terms (e.g., a task is either completed successfully or not) (van Mastrigt et al., 2023), task error could also be viewed on a continuous scale that measures the degree of deviation from a desired outcome (Kim et al., 2019). In a goal directed movement, task error could be quantified by how far off the result of a movement or action is from the intended target (Taylor et al., 2014). The basis for the reward prediction error is a reward consequence which is a subjective measure of usefulness, modulated by dopamine through reinforcement learning (Izawa & Shadmehr, 2011). Reward learning serves as a basis for strategic corrections and doesn't influence the internal model (Spampinato & Celnik, 2021; Taylor et al., 2014). Task errors are believed to elicit the use of explicit compensatory strategies, such as re-aiming, which allow individuals to consciously adjust their movements to achieve task goals (Miyamoto et al., 2020; Taylor & Ivry, 2011). This form of explicit learning is considered to be flexible and under volitional control, meaning it can be engaged or disengaged depending on its utility (McDougle et al., 2016). Experimental studies have explored explicit strategy use by either instructing participants to aim at specific locations or by asking them to generate and report their own aiming strategies (Mazzoni & Krakauer, 2006a; Miyamoto et al., 2020; Taylor et al., 2014).

1.1.4 Experimental Paradigms

A great deal of research on motor learning has traditionally focused on tasks involving arm movements such as reaching or pointing to a target. To study adaptation, these paradigms often introduce systematic perturbations that induce predictable changes in motor behavior (Izawa et al., 2008). Common experimental approaches include visuomotor rotations (Cunningham, 1989; Krakauer, 2009) and force perturbations (Shadmehr & Mussa-Ivaldi, 1994).

The visuomotor rotation paradigm is a widely used method in motor learning research to investigate how individuals adapt to altered visual feedback during movement. In this paradigm, participants perform reaching movements while viewing a cursor that represents their hand position on a screen. A rotation is applied to the cursor's movement relative to the actual hand trajectory. This visual perturbation causes a mismatch between the intended and perceived movement direction, leading to systematic errors. With practice, participants gradually adjust their movement direction to compensate for the rotated feedback, demonstrating the ability of the motor system to adapt to altered sensory conditions (Krakauer, 2009).

The force-field perturbation paradigm is a widely used approach in motor learning research to examine how the motor system adapts to dynamic changes in the environment. In this setup, participants perform reaching movements while holding the handle of a robotic manipulandum that applies velocity-dependent forces perpendicular to the direction

of movement, generating a curl field. These external forces initially disrupt the natural trajectory of the hand, leading to movement errors. With repeated exposure, individuals gradually adjust their motor output to successfully compensate for the perturbation (Shadmehr & Mussa-Ivaldi, 1994). Similar adaptive mechanisms have also been observed in response to altered inertial loads (Sainburg et al., 1999) and novel Coriolis forces (Lackner & Dizio, 1994).

While these paradigms have been instrumental in uncovering core principles of sensorimotor adaptation, they typically constrain movement to isolated joints or limbs and operate in environments with minimal ecological relevance. In contrast, whole-body motor learning paradigms offer a broader and more ecologically valid perspective (Babič et al., 2016, 2025). Unlike conventional tasks centred on discrete limb movements, these paradigms introduce realistic physical and biomechanical constraints, highlighting the critical roles of stability, coordination, and risk in shaping motor behaviour. Notably, they engage a more complex form of reward prediction error—one that extends beyond sensory discrepancies to encompass task-level outcomes directly linked to safety and survival, such as maintaining balance or avoiding injury. This shift in emphasis reflects a growing recognition that motor learning in natural environments often requires full-body coordination under dynamic and uncertain conditions, potentially engaging neural strategies that differ from those recruited in simplified, reductionist tasks.

1.1.4.1 Adaptation Process

When a perturbation is introduced in behavioural experiments, participants initially have a difficult time reaching for a target due to the changed environment. At the beginning, errors in movements are large and movement trajectories differ from those during the unperturbed movements. After some time, participants adapt, their errors decrease and they are able to reach the target. The adaptation process causes motor behaviour to result in smaller errors and smoother trajectories. This is partly due to a new formation of an internal representation of the relationship between motor commands and movement (D. W. Franklin et al., 2008). When the environment is reverted to the initial condition, a similar effect in the opposite direction is observed. Participants do not immediately return to their previous motor behaviour. Instead, they use newly learned motor commands that result in errors in movements. This is referred to as an after-effect. After-effects can be observed during catch trials, which are introduced during the adaptation period and participants are not aware of them. After-effects can be also observed during the de-adaptation trials, where the perturbation is similarly turned off as during the catch-trials, but with participants being informed about this changed condition. The only difference in these trials is the conscious awareness of the condition – knowledge of the perturbation being present or not. After-effects show that the participants learned their new behaviour and were not only adapting to perturbations with a feedback mechanism.

1.2 Thesis Purpose, Goals and Hypothesis

1.2.1 Purpose of the Dissertation

The purpose of this dissertation is to investigate the distinct contributions of sensorimotor and reward-based learning processes to human motor adaptation. Specifically, it aims to dissociate these two learning mechanisms, sensorimotor learning driven by sensory prediction errors and reward learning guided by reward prediction errors, and to characterize their individual roles and interactions in shaping motor behaviour. This research seeks to understand how each process contributes to adaptation when the environment undergoes abrupt changes, highlighting differences in implicit and explicit learning strategies. Furthermore, the dissertation explores how these learning processes operate in ecologically valid contexts by examining motor adaptation in whole-body tasks that incorporate realistic physical constraints and safety considerations. Additionally, the dissertation develops a computational framework that formalizes sensorimotor and reward-based learning processes to reproduce and predict observed patterns of human motor adaptation. Through a combination of behavioural experiments and computational modelling, the study aims to reveal how sensorimotor and reward learning integrate to support flexible, efficient, and safe motor control in dynamic real-world environments.

1.2.2 Goals of the Dissertation

To realize the purposes of this dissertation three main goals are set.

G1. Successfully dissociate sensorimotor and reward learning processes in the behavioural experiments.

G2. Observe the characteristics of sensorimotor and reward learning processes during the behavioural experiments separately and compare them.

G3 Determine how different error signals contribute to motor learning in a whole-body experiment that simulates an ecologically more relevant environment.

G4. With the help of the computational model show how sensorimotor and reward learning are combined to drive motor learning.

1.2.3 Hypothesis

In line with the purpose and goals of the dissertation we formulated the following hypothesis:

H1. There are two distinct processes of motor learning; reward learning and sensorimotor learning.

H2. Sensorimotor learning is an implicit process based on sensory consequence of movement and alters the internal model which produces a pronounced aftereffect and limits the ability of dealing with random changes in the environment.

H3. Reward learning is an explicit process based on reward consequence and does not produce aftereffects which allows for quick adaptation to randomly changing environment.

H4. A more complex and real-world motor task would require a combination of reward and sensorimotor learning for a quick and successful adaptation.

H5. An adaptive computational model that uses a control signal consisting of reward learning, sensorimotor learning and feedback controller in its learning algorithm would be able to model the processes of distinct reward and sensorimotor learning and replicate results observed in behavioural experiments.

1.3 Thesis Structure

This thesis is organized into six main chapters, each addressing a distinct aspect of motor learning and adaptation.

Following the introductory chapter, Chapters 2 and 3 present our work on interplay of sensorimotor and reward learning, with a focus on dissociating these learning processes based on error information. Both chapters investigate the same underlying principles but examine them through two distinct experimental paradigms and are therefore presented separately. Chapter 2 examines the contributions of sensorimotor and reward learning during experiments involving visuomotor perturbations and is based on the work on explicit strategies facilitating agile motor adaptations (Kunavar et al., 2023). Chapter 3 extends this investigation to force-field perturbations and is based on the work regarding distinct error signals and their impact on motor adaptation that is currently in submission (Kunavar et al., 2026).

Chapter 4 examines motor learning within an ecological context that requires balancing stability and movement efficiency. Through whole-body movement experiments, it demonstrates how sensorimotor and reward learning interact to enable safe and efficient adaptation. This work corresponds to the published paper Success-efficient/failure-safe strategy for hierarchical reinforcement motor learning (Babič et al., 2025) with shared first authorship.

Chapter 5 introduces a computational modelling framework that captures the interaction between sensorimotor and reward learning processes and replicates behavioural outcomes observed in the experimental data. The models presented in this chapter are based on a common underlying principle, with variations tailored to different experimental paradigms. The first part of this chapter describing iterative learning control model for visuomotor adaptation is part of the work on explicit strategies facilitating agile motor adaptations (Kunavar et al., 2023), while the second part of this chapter describing adaptive iterative learning control model for force-field adaptation is part of the manuscript on adaptive iterative learning control model integrating sensorimotor and reward learning, that is currently in submission (Kunavar & Babič, 2026).

Finally, Chapter 6 presents the overall conclusions of the thesis, summarizing key findings and their implications for future research in human motor learning and control.

Chapter 2

Explicit Learning Based on Reward Prediction Error Facilitates Agile Motor Adaptations

2.1 Introduction

To adapt our motions to changes in the environment our central nervous system mostly utilizes sensory prediction error and reward prediction error (Izawa & Shadmehr, 2011; Nikooyan & Ahmed, 2015; Tsay et al., 2022; Wu et al., 2022). The process of learning through sensory prediction error is known as sensorimotor adaptation, whereas learning that leverages reward prediction error is referred to as reward learning. The basis for the sensory prediction error is the sensory consequence of motion, observed by our sensory organs such as vision and proprioception. Sensorimotor adaptation is a process that allows alterations to the internal model either through a feedforward model, which maps motor commands to corresponding sensory consequences (Kawato, 1999; Shadmehr et al., 2010), or through policy learning (Hadjiosif et al., 2021). These alterations to the internal model occur without conscious awareness and can result in large and prolonged aftereffects, reflecting changes made to the internal model during prior adaptation (Kawato, 1999; Miyamoto et al., 2020). Reward prediction error is the difference between internal reward prediction and the actual observed reward (Sedaghat-Nejad et al., 2019; Wu et al., 2022). The basis for the reward prediction error is the reward consequence, a subjective measure of usefulness, modulated by dopamine through reinforcement learning (Izawa & Shadmehr, 2011; Schultz et al., 1997).

Motor learning can rely either on sensory prediction error, reward prediction error, or both. When sensory feedback is available, adaptation is predominantly driven by sensory prediction error (Izawa & Shadmehr, 2011), while in the case of limited sensory information, motor learning is driven by reward prediction error (Holland et al., 2018; Nikooyan & Ahmed, 2015). Studies have demonstrated that relying on either type of prediction error can yield similar levels of motor adaptation (Izawa & Shadmehr, 2011; Palidis et al., 2019).

For example, when hitting the ball in tennis, you can observe the sensory consequences of your swing, like the position of your arm, through vision or proprioception during the

swing. The sensory prediction error arises when the sensory information does not align with your expected motion. This error results from the neural noise or external factors like a changed racket weight. Sensorimotor adaptation accounts for this error by making alterations to the internal model without our conscious awareness. Furthermore, success or failure in hitting the ball translates to a reward consequence, expressed as a reward prediction error. A successful hit means you will likely repeat the same motion; however, a miss might lead you to consciously adjust your swing. Reward prediction error proves particularly useful when dealing with a changing environment, such as strong wind during a tennis match. If there is a wind, you might adjust your swing towards the opposite direction to compensate for the wind resistance. In the real world, there are many such parameters that attribute to the changing environment and the human motor control system is especially good in dealing with such randomly changing perturbations, which affect our movements.

Based on previous studies showing that sensorimotor learning can result in large and prolonged aftereffects reflecting the changes to the internal model during previous adaptation, we hypothesize that learning based on sensory prediction error might limit the ability of participants to deal with randomly changing perturbations. Specifically, sensorimotor adaptation to a particular perturbation during an earlier task will not help participants in the task where perturbations vary randomly. On the other hand, participants learning from reward consequences would use explicit learning strategies that would not produce aftereffects. We hypothesize that learning based on reward prediction error would allow participants to adapt quicker to randomly changing perturbations. To test these hypotheses, we used a visuomotor paradigm (Cunningham, 1989; Imamizu et al., 1995; Krakauer, 2009) where the goal of the motor task is to hit a designated target region with a visual cursor controlled by the motion of a hand. During the learning phase, an angular rotation is imposed on the cursor and participants have to learn to counter the perturbation by moving the hand in the opposite direction (Bond & Taylor, 2015). Moreover, to differentiate between reward and sensory prediction errors, we modified the task as in (Izawa & Shadmehr, 2011; Palidis et al., 2019) by presenting participants either the information on sensory consequences of their movement, signed end-point error or just binary reward.

2.2 Results

Participants were asked to adapt to visuomotor rotation of 30° in the clockwise or counter-clockwise direction while reaching towards a target with their right hand (Figure 2.1A). Before each trial, perturbation indicators in the form of visual cues were presented to inform the participants of the presence and direction of the perturbation. Their task was to hit the target when it appeared on the screen in front of them. After each motion, a cross mark showed the location where the hand passed the distance of the target. Additionally, the target changed its colour to green if it was hit or to red if it was missed. There are two types of motor errors present during this kind of motion: sensory prediction error which is defined as the difference between the expected and the observed position of the hand during the motion (represented as the location of the cursor), and reward

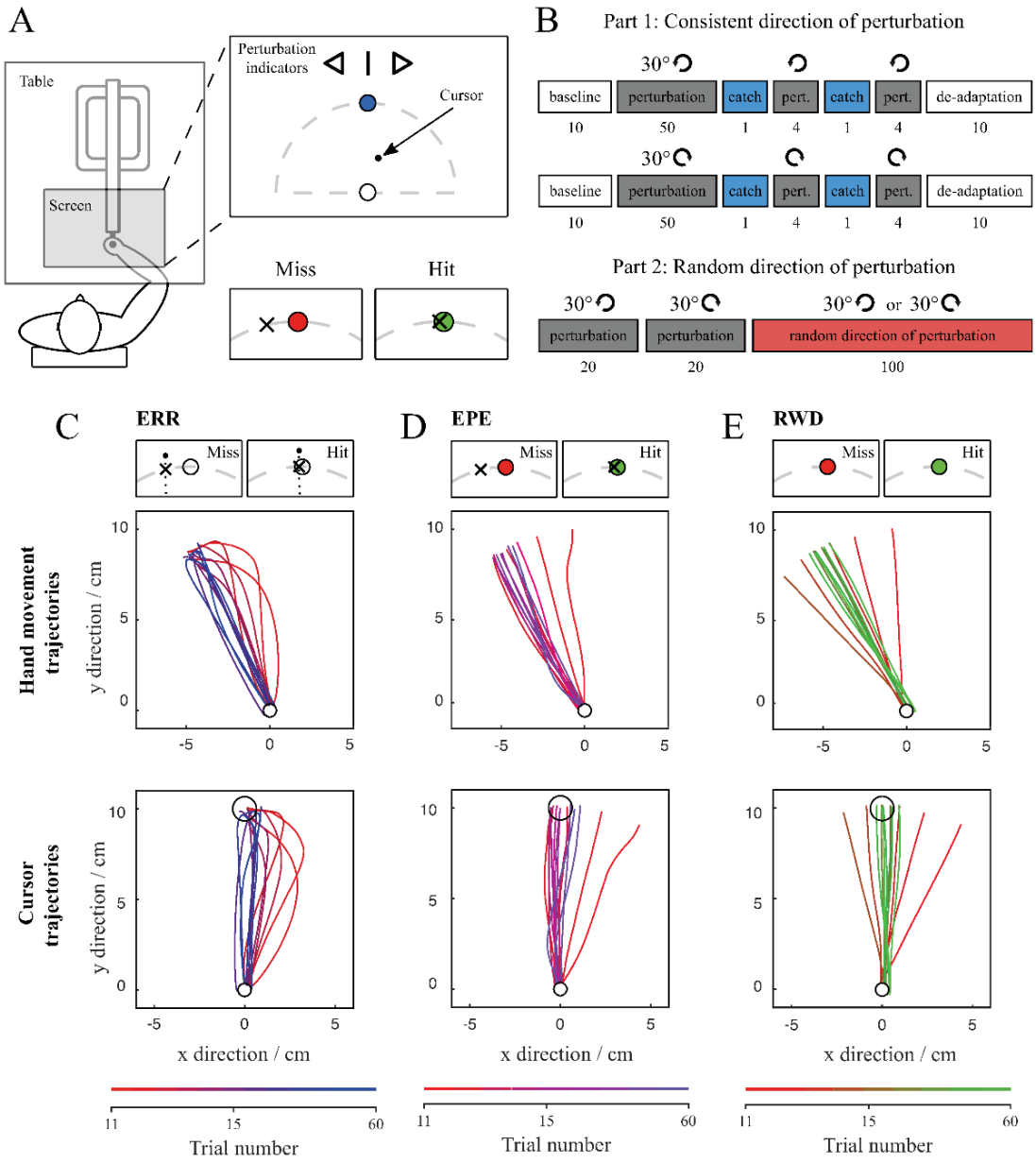


Figure 2.1: Experiment (visuomotor rotation). (A) Experimental setup. Participant is holding a handle of the Haptic Master covered with a screen. Top right figure shows a screen with a start position, target, cursor, and perturbation indicators. Bottom right figure shows indicators of a missed or hit trial. (B) Experimental protocol. The experiment is divided into two parts: in the first part the direction of perturbation is consistent whereas in the second part the direction of perturbation is random. White blocks represent trials without any perturbation. Grey blocks represent trials with a consistent perturbation in one direction, while the red block represents trials where the direction of perturbation is random. Blue blocks represent trials when the perturbation was omitted (catch trials). Arrows show the direction of perturbation and numbers below the blocks denote the number of trials per specific block. The order of the perturbation direction (clockwise or counter clockwise) was counterbalanced across participants. (C–E) Example cursor and hand trajectories for the three experimental groups: full visual error (ERR), end-point error (EPE), and binary reward (RWD). The trajectories are shown for three representative

participants for the initial five perturbed trials followed by every sixth trial. The progression of the trials is represented with the transition of the red into blue colour for the ERR group, red into violet colour for the EPE group and red into green colour for the RWD group.

prediction error which is defined as the difference between the expected and the observed success of the motion. To dissociate these two types of errors, we assigned participants into three groups based on the error information that was provided to them. First group received full visual feedback to study the influence of the sensory prediction errors throughout the reaching movement (ERR group). The second group received end-point error information at the end of the movement (EPE group). Compared to the binary reward, the end-point error in the EPE group was meant to emulate a more realistic situation in our daily lives, where we usually know the size and direction of our mistakes. It also avoided the very extensive exploration needed in the case of a binary reward. The third group received only binary reward feedback at the end of the movement (RWD group) as the reward prediction error. The reward feedback was given immediately after a trial was completed. Participants first learned to adapt to clockwise and counter clockwise perturbations separately and then proceeded with the randomly changing direction of the perturbation (Figure 2.1B).

During the initial baseline movements, when there was no perturbation, all the ERR, EPE and RWD groups exhibited comparable movements with nearly straight-line hand and cursor trajectories and the Cursor path area values close to zero. The average Cursor path area of baseline trials was -0.22 ± 3.65 cm^2 for the ERR group, 1.18 ± 4.56 cm^2 for the EPE group and 0.80 ± 5.28 cm^2 for the RWD group. Statistical analysis showed no significant difference between the groups ($F(23) = 0.20$, $p = .817$). Average Cursor angle of baseline trials was -1.45 ± 7.28° for the ERR group, 2.05 ± 6.26° for the EPE group and 0.80 ± 5.28° for the RWD group, with no significant difference between the groups ($F(23) = 0.63$, $p = .542$).

2.2.1 Part 1: Consistent Direction of Perturbation

The aim of the first part of the experiment was to investigate the adaptation to visuomotor perturbation and to explore the possible differences in motor learning between the ERR, EPE and the RWD groups. All participants received the information about the direction of the perturbation in the form of visual cues before the start of each trial. ANOVA analysis showed groups differed in the time needed to perform the movement ($F(23) = 9.28$, $p = .001$). Post-hoc t-tests further showed that on average, the ERR group needed more time to perform the movements (1.09 ± 0.31 s) compared to the EPE group (0.63 ± 0.16 s; $t(14) = 3.70$, $p = .002$) and RWD group (0.70 ± 0.19 s; $t(14) = 3.05$, $p = .007$).

When the perturbation was first introduced, the hand movement trajectories remained straight while the cursor trajectories shifted in the same direction as the perturbation. Figure 2.1C-E shows the hand movement trajectories and cursor trajectories for the ERR, EPE and RWD groups. It should be noted that the EPE and RWD groups did not see the actual cursor during the experiment. In the ERR group, initial cursor trajectories were curved and participants still managed to hit the target, presumably by using visual

feedback during the movement. After a few trials, cursor trajectories became notably less curved as participants proceeded to learn. In contrast to the ERR group, the initial cursor trajectories of the EPE and RWD groups were almost straight but rotated and with a larger end-point error. During the adaptation, participants kept the cursor trajectories straight, but changed the direction of movement which decreased the end-point error. While participants in the EPE group saw the end-point error and could change the direction of the movement accordingly, the participants in the RWD group had to explore the movement space to find a movement where they hit the target.

Participants in all groups successfully adapted to visuomotor rotation, which can be seen as a decrease in Cursor path area during perturbation trials (Figure 2.2A). In effect, cursor trajectories became similar to the baseline cursor trajectories with Cursor path area and Cursor angle coming close to zero for all groups.

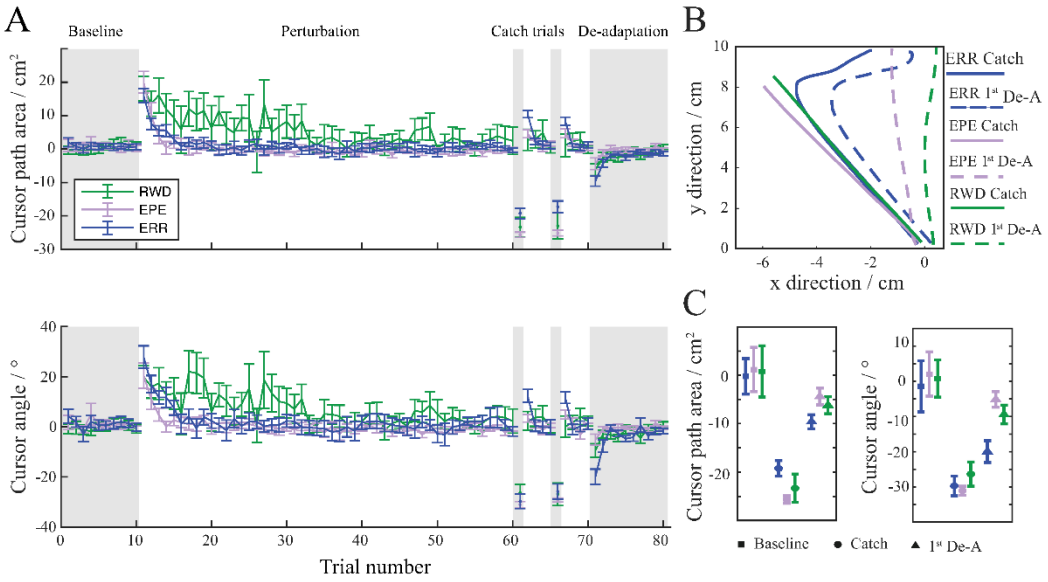


Figure 2.2: Movements during consistent direction of perturbation. (A) Mean values of Cursor path area and Cursor angle combined for both directions of perturbation showing the progression of learning. Cursor path area and Cursor angle for the ERR group are shown in blue, for the EPE group in violet and for the RWD group in green. Error bars show standard error of the mean. Shaded areas represent trials without perturbation and non-shaded areas represent trials with perturbation. (B) Example trajectories of catch and de-adaptation trials. Because there is no perturbation for these trials, cursor trajectories correspond to the hand movement trajectories. Trajectories from the ERR group are shown in blue, for the EPE group in violet and for the RWD group in green. (C) Comparison of aftereffects between the ERR, EPE and RWD groups. Average Cursor path area and Cursor angle for baseline, catch and de-adaptation trials. Error bars show standard deviation.

To assess learning, we looked at the Cursor path area of catch trials when the perturbation was absent. Absolute Cursor path area values of catch trials were comparable between all groups. There was no significant difference between groups for the first ($F(47) = 2.73$, $p = .076$) as well as the second catch trial ($F(47) = 2.58$, $p = .087$). Since the ERR

group was provided with sensory information and was able to correct their movements during a single movement, we additionally looked at the initial part of the movement (Cursor angle), before the contribution of the feedback. During the catch trials, the initial part of the arm reaching movements was similar for participants in all groups and there was no difference between the corresponding Cursor angle for neither the first catch trial ($F(47) = 0.82$, $p = .448$) nor for the second catch trial ($F(47) = 0.29$, $p = .753$).

To assess the after-effects of learning, we looked at the 1st de-adaptation trial. The perturbation was removed for de-adaptation trials. Contrary to the catch trials, participants were informed of the lack of perturbation during de-adaptation trials. However, visual information during the trial was the same as in previous trials and there was no difference between catch trials and de-adaptation trials, except for the difference in visual cues. There was an apparent difference in Cursor angle between the 1st de-adaptation trials of all groups ($F(46) = 7.99$, $p = .001$). The average value of Cursor angle for the 1st de-adaptation trial of group ERR was significantly larger (in absolute terms) than the Cursor angle for the 1st de-adaptation trial of the group EPE ($t(30) = 3.89$, $p = .001$) and RWD ($t(29) = 2.52$, $p = .028$). Participants in the ERR group had a large aftereffect with curved trajectories similar to those during the catch trials ($t(15) = 1.39$, $p = .188$) (Figure 2.2B-C). Since there was no perturbation present during the 1st de-adaptation trials and catch trials, cursor trajectories correspond to hand movement trajectories. The average value of Cursor angle for the 1st de-adaptation trial was significantly larger (in absolute terms) than the average value of Cursor angle for the baseline trials ($t(15) = 6.96$, $p < .001$). On the other hand, participants in the groups EPE and RWD exhibited smaller aftereffect during the 1st de-adaptation trial ($t(15) = 10.53$, $p < .001$ for the EPE group and $t(14) = 3.05$, $p = .008$ for the RWD group). Their trajectories were closer to those during the baseline trials (Figure 2.2B-C), however, the Cursor angles of 1st de-adaptation trials were still significantly different from Cursor angles of baseline trials ($t(15) = 3.07$, $p = .008$ for the EPE group and $t(13) = 3.10$, $p = .008$ for the RWD group).

2.2.2 Part 2: Alternating Direction of Perturbation

In the second part of the experiment, participants were presented with a randomly changing direction of the perturbation to investigate how the two different prediction errors affect the ability to successfully perform the motion when the direction of perturbation switches from trial to trial. Participants first performed 20 trials with clockwise perturbation and 20 trials with counter clockwise perturbation (Figure 2.3A) and then proceeded to the random direction of perturbation (Figure 2.3B). The sequence of the trials was random; however, all participants received the information about the direction of the perturbation in the form of visual cues before the start of each trial. We wanted to see if participants can successfully adapt their movements to two opposing perturbations based on the advance information about perturbation. The cursor trajectories of participants in the ERR group were similar to the initial cursor trajectories from the first part of the experiment when participants first experienced the perturbation and had not yet adapted their motor behaviour. Their cursor trajectories were curved to the left when there was a counter clockwise rotation (light blue on Figure 2.3D) and to the right when there was a clockwise rotation (dark blue on Figure 2.3D). In stark contrast, participants in the EPE

and RWD groups could easily adapt to the perturbation that was randomly changing in direction. Their cursor trajectories were almost straight (light and dark violet lines for the EPE group and light and dark green lines for the RWD group on Figure 2.3E-F) with no apparent difference between the two directions of perturbation.

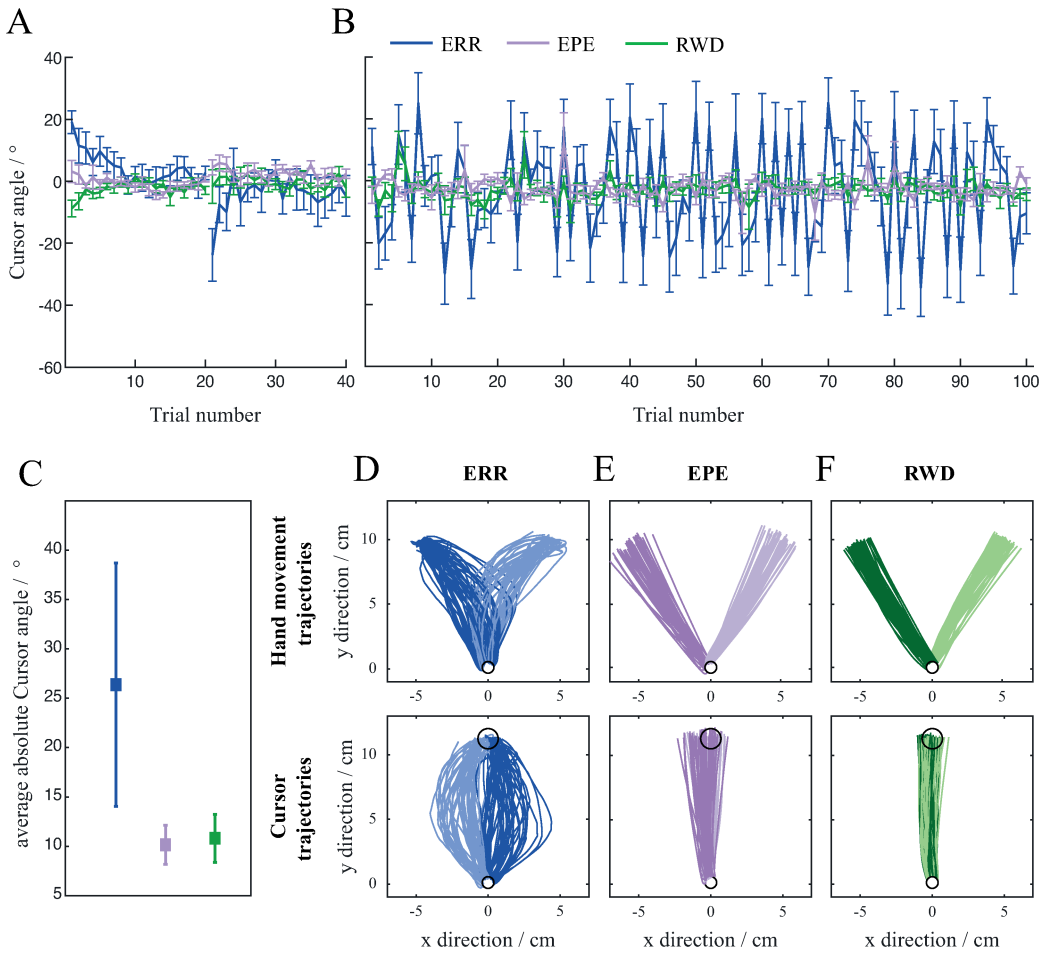


Figure 2.3: Movements during random direction of perturbation. Mean Cursor angle during short block of consistent direction of perturbation (A) and random direction of perturbation (B). Mean Cursor angle is presented as a function of trials. Cursor angle for the ERR group is shown in blue, for the EPE group in violet and for the RWD group in green. Error bars show standard error of mean. (C) Cursor trajectories of a participant from the ERR group. Light blue denotes trajectories with perturbation in the left direction and dark blue denotes trajectories with perturbation in the right direction. (D) Cursor trajectories of a participant from the EPE group. Light violet denotes trajectories with perturbation in the left direction and dark violet denotes trajectories with perturbation in the right direction. (E) Cursor trajectories of a participant from RWD group. Light green denotes trajectories with perturbation in the left direction and dark green denotes trajectories with perturbation in the right direction.

Even though participants in all groups had previously learned to counteract the perturbation in opposite directions, only the EPE and RWD groups were able to deal with the perturbation in random direction. Participants in the ERR group made large errors

and were not able to adapt to changing direction of perturbation, which is reflected in higher Cursor angle values. On the other hand, participants in the EPE and RWD groups made smaller errors and were able to adapt to the changing direction of perturbation, which is reflected in smaller Cursor angle values that are closer to zero. There was a significant difference in mean Cursor angle between the groups ($F(23) = 12.49$, $p < .001$). Mean Cursor angle of the ERR group was significantly larger than the mean Cursor angle of the EPE group ($t(14) = 3.67$, $p < .001$) as well as the RWD group ($t(14) = 3.50$, $p = .001$). There was also a difference in inter participant variance between groups ($K^2(2) = 25.04$, $p < .001$). Variance in ERR group was higher compared to the variance in the EPE ($F(1) = 25.83$, $p < .001$) and the RWD groups ($F(1) = 38.93$, $p < .001$).

During the 100 trials with random direction of perturbation, there was no additional adaptation to the alternating perturbations. In both groups, Cursor angle of the first 10 trials was not significantly different as Cursor angle of the last 10 trials ($t(14) = 0.08$, $p = .936$, $t(14) = 0.50$, $p = .626$, $t(14) = 1.09$, $p = .293$ for the ERR, EPE and RWD group respectively).

2.3 Discussion

To compare and analyse learning based on sensory prediction error versus learning based on reward prediction error, we used a visuomotor paradigm where participants performed arm reaching movements and were presented either with the information on sensory consequences of their movement or reward consequence of their movement.

The aim of the first part of the experiment was to investigate the adaptation to visuomotor perturbations and to explore the possible differences between the sensorimotor adaptation, end-point error learning and binary reward learning. Participants could successfully adapt to visuomotor rotation in all cases, i.e., relying either on sensory or on reward prediction error. There were similar levels of motor adaptation, observed through comparable Cursor path area and Cursor angle during adaptation, and similar catch-trial aftereffects. This is consistent with the findings of previous studies (Izawa & Shadmehr, 2011; Palidis et al., 2019) showing that motor learning can rely on any form of prediction error and still lead to similar levels of motor adaptation.

De-adaptation trials revealed differences in the aftereffects. As expected from previous works (Miyamoto et al., 2020; Taylor et al., 2014), learning based on sensory prediction error results in large and prolonged aftereffects. This was observed in the process of de-adaptation where there was an apparent difference between the base trials and the 1st de-adaptation trials. These apparent aftereffects suggest that participants in the ERR group used sensorimotor adaptation to alter their internal model during the adaptation. On the other hand, explicit learning based on reward prediction error did not produce aftereffects as with sensorimotor adaptation. Participants did not exhibit any aftereffect during the 1st de-adaptation trial and their trajectories resembled baseline trials. This coincides with previous work, where they showed that reward information alone failed to induce sensorimotor recalibration (Nikooyan & Ahmed, 2015; Tsay et al., 2022). However, this might only apply to instances where only one target is used in the experimental setup, since it was shown that learning based on end-point error produces higher aftereffects than

observed in our study in cases where multiple targets are presented (Taylor et al., 2014). This might be because explicit strategies are harder to implement in cases of multiple targets.

To investigate how the two different prediction errors affect the ability to successfully perform the motion when the direction of perturbation switches from trial to trial, participants were presented with a randomly changing direction of perturbation. Even though participants relying on sensory prediction error had previously learned to counteract the perturbation in both directions, they could not deal with the perturbation in random direction. This suggests that they could not make timely changes to the internal model and distinguish between motor commands obtained for a specific perturbation. This is in line with previous research, which showed that sensorimotor adaptation of distinctive perturbations does not allow for simple switching between motor commands during randomly changing perturbations when they are distinguished by visual cues (Karniel & Mussa-Ivaldi, 2002; Wada et al., 2003). On the other hand, participants relying on reward prediction error (either end-point error or binary reward) were able to easily adapt to the random direction of perturbation. Similar results were found in studies that showed dual adaptation was possible primarily through explicit strategies (Hegele & Heuer, 2010; Schween et al., 2018). Interestingly, the analysis of the inter-participant variance showed that the EPE and RWD groups had smaller variance between participants compared to the ERR group. This might be because the explicit information on the direction of perturbation with the cues encouraged explicit strategies even in the ERR group for some participants. Explicit motor learning can be considered as a conscious process that serves as a basis for strategic corrections and doesn't influence the internal model (Spampinato & Celnik, 2021; Taylor et al., 2014). This suggests that explicit learning allows for quicker adaptation of motor behaviour. However, this only applies in cases when cues are indirect, meaning they are not part of the dynamic state of the body or manipulated object. In cases where cues are direct, such as the use of separate hands (Nozaki et al., 2006; Schween et al., 2019), movement locations (Howard et al., 2013), visual workspace locations (Forano et al., 2021), or movement plans (Hirashima & Nozaki, 2012; Howard et al., 2012, 2015; Sheahan et al., 2016), sensorimotor adaptation allows for switching between motor commands during randomly changing perturbations.

We argue that only sensorimotor adaptation affects internal model which produces an aftereffect in cases when a person is aware of the lack of perturbation. The changes to the internal model that occurred are not consciously perceived so the person does not know that their hand is moving in another direction. Therefore, when the perturbation is removed, they cannot instantaneously make changes to the internal model in order to compensate for the change in perturbation. On the other hand, explicit reward learning does not affect the internal model, but rather the participants consciously adapt the direction in which they aim. Therefore, they are aware of the changes to their movement and have a better understanding of where the real position of the hand actually is. When they are informed of the change in perturbation, they can thus consciously change the direction of the movement towards the right target.

Our results support that the sensorimotor adaptation and explicit reward learning work through distinct adaptation processes. This suggest that only sensorimotor adaptation

makes changes to the internal model, while explicit reward learning uses strategies that do not alter internal model and thus do not produce aftereffects. Additionally, we could show that when humans learn motor tasks, they use both learning processes to successfully adapt to the new environment. This contributes to the understanding of human motor adaptation and the importance of separately looking at the sensorimotor adaptation and explicit reward learning, as well as the importance of taking into account both learning processes since humans tend to use a combination of both.

2.4 Methods

2.4.1 Participants

Twenty-four right-handed volunteers (6 females and 18 males) participated in the study. They were divided in three groups, the ERR group (n = 8, 2 females and 6 males), the EPE group (n = 8, 2 females and 6 males) and the RWD group (n = 8, 2 females and 6 males). Their average age was 27.34 ± 6.61 years, height 178.00 cm ± 10.25 cm and body mass 72.52 ± 15.85 kg. Prior to their participation, the participants were informed about the course of the experiment and signed a written consent approved by the Slovenian National Medical Committee (No. 339/2017/7). We have pseudonymized all the data obtained from experiments. All experimental protocols were approved by the National Medical Ethics Committee Slovenia (No. 339/2017/7) and the methods were carried out in accordance with the relevant guidelines.

2.4.2 Experiment

Participants were seated in a chair in front of a table with an integrated screen (Figure 2.1A). They held the handle of a haptic robot (Haptic Master MK2, Moog, Nieuw-Vennep, The Netherlands) that was located underneath the table and not visible to the participants. The participants were asked to perform a series of arm reaching movements. The position of the participant's hand was displayed on the screen as a cursor.

The ERR group was provided with both types of prediction errors by showing them the cursor during the whole reaching task. Here we assumed that the participants primarily rely on the sensory prediction error since in the presence of sensory information participants predominantly disregard the information of the task success (Izawa & Shadmehr, 2011). The EPE and RWD groups had the cursor hidden to them during the motion. The EPE group received the information on where they passed the distance of the target (end-point error) and whether the target was hit or missed. The RWD group only received the information whether the target was hit or missed. Since there was no visual feedback during the movement, information on sensory prediction error was minimized and participants were only able to rely on information regarding the task success (reward prediction error).

There were two types of trials: unperturbed and perturbed. During the unperturbed trials, the motion of the cursor was aligned with the actual movement of the hand while during the perturbed trials the cursor underwent a rotation of 30° in either clockwise (right direction of perturbation) or counter clockwise direction (left direction of perturbation). In

effect, when the hand moved straight ahead, the cursor moved either 30° towards the right or towards the left.

A trial started with the Haptic Master positioning the handle to a start position displayed as a white circle with a diameter of 0.8 cm. The target towards which the participants had to perform the arm movement was a blue circle with a diameter of 1.4 cm which was located 10 cm away from the start position. On the top of the screen, there was a graphical perturbation indicator that displayed the direction of the perturbation in the form of visual cues (straight line for unperturbed trials, right arrow for right direction and left arrow for left direction of perturbation). The visual cues were used because we wanted to measure feedforward learning and not feedback response. Visual cues are used for all trials and, with the exception of the catch trials, they always reflect the true perturbation. Participants were told that there would be trials without perturbation and trials with perturbation where the position on the screen would not accurately reflect the position of their hand. They were informed that the visual cues will be shown before each trial. They were only given the explanation that visual cues show the direction of perturbation, however no further information on the type of perturbation was given. Participants were informed about the removal of the perturbation using a corresponding visual cue. Furthermore, they were reminded to pay attention to the perturbation indicators a few trials before the start of the de-adaptation trials.

2.4.3 Experimental Protocol

The experiment was divided into two consecutive parts as shown on Figure 2.1B. In the first part, the direction of perturbation was consistent, while in the second part, the direction of perturbation was random.

2.4.3.1 Consistent Direction of Perturbation

The first part with the consistent direction of perturbation was meant for participants to adapt to a new perturbation. Participants first learned to adapt to one perturbation and then proceeded with the other direction of perturbation. Half of the participants started with clockwise rotation and the other half started with the counter clockwise rotation.

Participants first performed 10 baseline trials without perturbation, followed by 50 trials with perturbation in a single direction. To assess the adaptation to perturbation, we included 2 catch trials near the end of the perturbed trials. Each catch trial was followed by 4 perturbed trials. The first part concluded with 10 de-adaptation trials without perturbation (upper part of Figure 2.1B).

2.4.3.2 Alternating Direction of Perturbation

The second part of the experiment was designed to measure adaptation to the randomly changing direction of the perturbation. It started with 20 perturbed trials with clockwise rotation followed by 20 perturbed trials with the counter clockwise rotation. Afterwards, the participants had to perform a randomly ordered set of 100 trials with either the clockwise or counter clockwise rotation (lower part of Figure 2.1B). The pattern of the perturbation directions was consistent for all participants.

2.4.4 Data Analysis

Position of the hand during the arm reaching motion was recorded by the haptic robot with a frequency of 30 Hz. Motion data was filtered using a 4^{th} order low-pass Butterworth filter with 10 Hz cut-off frequency. In addition to the hand movement trajectories, we also calculated the cursor trajectories based on the visuomotor rotation that was applied in each trial. To observe motor learning, cursor trajectories were segmented into individual trials where the start of the trial was defined as the moment when the movement trajectory crossed the 0.4 cm target distance relative to the starting position and the end of the trial was defined as the moment when movement trajectory crossed the 10 cm distance relative to the starting position. To assess the effects of perturbation on the motion and the adaptation to perturbation over the trials, we calculated Cursor path area for each trial. Cursor path area is defined as the area between the cursor trajectory and a straight line between the start position and the target and takes into account the whole motion (Burdet et al., 2001). In cases where the target is not hit the Cursor path area is defined as the area between the cursor trajectory, the straight line between the start position and the target and a circle around the start area with a diameter of 10 cm. To assess the adaptation without the contribution of sensory feedback, we calculated Cursor angle (Day et al., 2016) defined as an angle between the y-axis of the coordinate frame and the vector between the cursor position at the beginning of the motion when the velocity reached 0.025 m/s and the cursor position 160 ms after the beginning of the motion. For the first part of the experiment, both the Cursor path area and Cursor angle during the counter clockwise rotation were multiplied with -1, so that the values were comparable to those obtained during clockwise rotation and could be averaged.

To compare ERR, EPE and RWD groups, we compared average trial times of individual participants, Cursor path area and Cursor angle values of catch trials, Cursor angle values of 1^{st} de-adaptation trials, and average Cursor angle values of individual participants for trials during random direction of perturbation between the three groups. ANOVA was used for all comparisons between the groups and additional post hoc t-tests with Bonferroni correction were conducted to determine the significant differences between the specific groups. Paired t-test was used to compare average Cursor angle values of the first 10 trials to the average Cursor angle values of the last 10 trials of each individual participant. To test equality of variances we used Bartlett test and additional post hoc F-tests of equality of variances with Bonferroni correction.

Chapter 3

Distinct Error Signals and Their Impact on Motor Adaptation: An Examination Through Force-Field Perturbation and Contextual Switching

3.1 Introduction

Motor adaptation is a fundamental process through which the human motor system adjusts to new conditions, ensuring our ability to perform movements accurately in novel or changing environments (Krakauer et al., 2019; Shadmehr et al., 2010; Wolpert et al., 2011). It reflects the brain's ability to learn from errors and to recalibrate motor commands in response to changing environmental dynamics (Kim et al., 2021; Krakauer & Mazzoni, 2011). It is believed that the motor system utilizes two distinct types of error signals to adapt our motions, a task error and a sensorimotor error (Albert et al., 2022; L. Leow et al., 2018; Tsay et al., 2022). Sensorimotor error is a difference between predicted sensory outcome and actual sensory outcome of movement (Kawato, 1999; Tseng et al., 2007; Wolpert et al., 1998; Wolpert & Flanagan, 2001). Task errors are discrepancies between predicted and actual reward and are presented in a goal directed movement as a failure to achieve an internally determined task goal (L.-A. Leow et al., 2020; Schultz, 2017). While some tasks might be assessed in binary terms (e.g., a task is either completed successfully or not) (van Mastrigt et al., 2023), task error could also be viewed on a continuous scale that measures the degree of deviation from a desired outcome (Kim et al., 2019). In a goal directed movement, task error could be quantified by how far off the result of a movement or action is from the intended target (Taylor et al., 2014). Studies with visuomotor perturbation have demonstrated that relying on either type of error can yield similar levels of motor adaptation (Izawa & Shadmehr, 2011; Palidis et al., 2019; Therrien et al., 2016). However, different types of error have distinct impacts on generalization, de-adaptation, and memory retention (Abe et al., 2011; Izawa & Shadmehr, 2011; Kunavar et al., 2023; K. Lee et al., 2018; Therrien et al., 2016).

The ability to foresee and adjust for potential errors in movement relies on the effective formation and selection of relevant motor memories (Česonis & Franklin, 2022; Forano et al., 2021; Oh & Schweighofer, 2019). The sensorimotor system's ability to adapt simultaneously to different disturbances depends on the presence of contextual cues, which are essential for recognizing context-dependent changes and enabling the system to differentiate between varied environments or demands, thereby applying the most appropriate motor response (Forano et al., 2021). Contextual switching studies showed that dual adaptation is effective when cues are direct - they pertain to the dynamic state of the body or manipulated object, such as different visual or physical workspace locations or distinct lead-in or follow-through movements (Forano & Franklin, 2020; Howard et al., 2012; Nozaki et al., 2006). In cases when cues are indirect, such as colour or arrows, simultaneous adaptation to different disturbances can lead to some type of interference (Howard et al., 2013; Karniel & Mussa-Ivaldi, 2002; 'Shadmehr et al., 1994). It was previously shown that adaptation based on direct cues is facilitated by implicit learning, while adaptation based on indirect cues is facilitated by explicit learning such as explicit strategy to aim in a different direction(Forano et al., 2021).

Simultaneous adaptation to different disturbances also depends on the type of error signal, since distinct error signals are thought to differently engage implicit and explicit learning. Sensorimotor error is associated with implicit learning, while task error is primarily associated with explicit learning (Bond & Taylor, 2015; Taylor et al., 2014; Taylor & Ivry, 2011; Tsay et al., 2022). It was recently shown that dual adaptation based on indirect cues was possible when error signal was in the form of task error, either end-point or binary error, and not when it was in the form of sensorimotor error (Kunavar et al., 2023). However, this has only been previously investigated in a visuomotor rotation task. Moreover, explicit strategy implementation may be reduced in tasks involving multiple targets due to limitations in conscious processing and working memory (Forano et al., 2021; Taylor et al., 2014). We were interested in whether similar effects of error type on contextual switching would be observed in our force-field perturbation study. Based on previous studies we would expect the participants to successfully learn to adapt to force perturbation regardless of the type of error information. However, only participants who received the task error would be able to adapt to switching perturbations.

To test this, we used a force-field perturbation paradigm (Gandolfo et al., 1996; Shadmehr & Mussa-Ivaldi, 1994), where the goal of the motor task was to hit a designated target region with a visual cursor controlled by the motion of a hand. During the learning phase, a force-field perturbation was imposed on the handle of the haptic robot and participants had to learn to counter the perturbation. To differentiate between error signals, we divided participants into three groups based on the information that was available to them. Participants received either the information on sensory consequences of their movement, signed end-point error or binary reward (Izawa & Shadmehr, 2011; Palidis et al., 2019). Experimental protocol was adapted from (Kunavar et al., 2023) and included a part where perturbation was consistent to measure adaptation and a part where direction of perturbation was randomly changing to test rapid adaptations to switching perturbations.

3.2 Results

Participants were asked to adapt to curl force-field perturbations while reaching towards a target with their right hand. Their task was to hit the target when it appeared on the screen in front of them. To distinguish between different types of errors, we assigned participants to three groups based on the error information provided to them (Figure 3.1A).

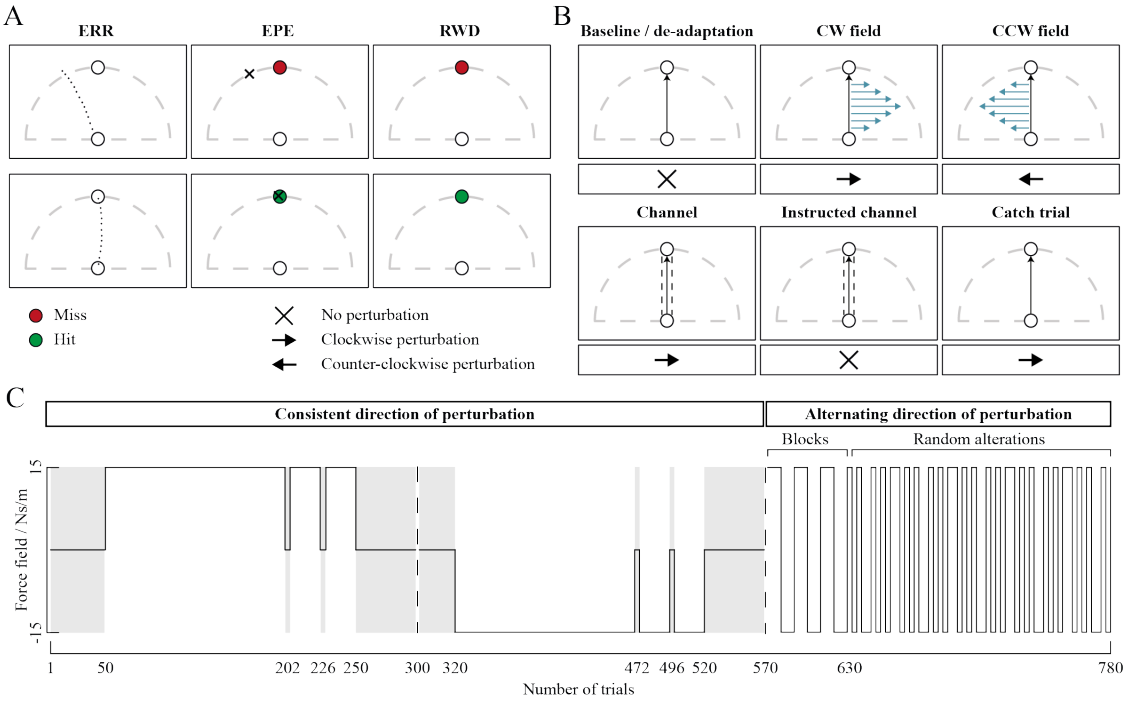


Figure 3.1: Experiment (force field perturbation). (A) Experimental groups. The ERR group saw the cursor during the whole reaching task. The EPE group received only the information on where they passed the distance of the target (end-point error) and whether the target was hit or missed. The RWD group only received the information whether the target was hit or missed; green for hit and red for miss. (B) Experimental conditions. Trials were either unperturbed, perturbed in the clockwise (CW) or counterclockwise (CCW) direction or confined to a simulated channel. Prior to each movement a visual cue was provided to the participants indicating the direction of the force field. There were two cases of incorrect visual cues: catch trials and channel trials. In catch trials, the perturbation was absent, but visual cues still indicated perturbation. In normal cahnnel trials, the cue indicated perturbation, but the movement was confined to a simulated mechanical channel. In instructed channel trials, the movement was confined to a simulated mechanical channel, but the cue indicated no perturbation. (C) Experimental protocol. The experiment is divided into two parts: in the first part the direction of perturbation is consistent whereas in the second part the direction of perturbation is random. The order of the perturbation direction (clockwise or counter clockwise) was counterbalanced across participants. Grey shaded blocks represent parts without perturbation and non-shaded blocks represent parts with either clockwise or counterclockwise perturbation.

The first group received full visual feedback throughout the reaching movement (ERR group), the second group received only end-point error information (EPE group), and the third group received only binary information (RWD group). The movements were either unperturbed, perturbed (by either a clockwise or counter-clockwise force field), or confined to a simulated channel (Figure 3.1B). Before each trial, perturbation indicators in the form of visual cues were presented to inform the participants of the presence and direction of the perturbation. Within each group, half of the participants reached towards 3 targets while the other half reached towards one target. Participants first learned to adapt to clockwise and counter-clockwise perturbations separately and then proceeded with the alternating direction of the perturbation (Figure 3.1C). During alternating directions of perturbation, first there were blocks of alternating force-field perturbations, followed by randomly alternating force-field perturbations. To evaluate learning, we calculated kinematic error in the middle of the movement (4 to 6 cm distance from the start target) and the end-point error at the end of the movement for each trial. Additionally, we calculated compensatory force measured during the channel trials. Figure 3.2 shows kinematic error throughout the experiment with a single target location and Figure 3.3 shows kinematic error throughout the experiment with three target locations.

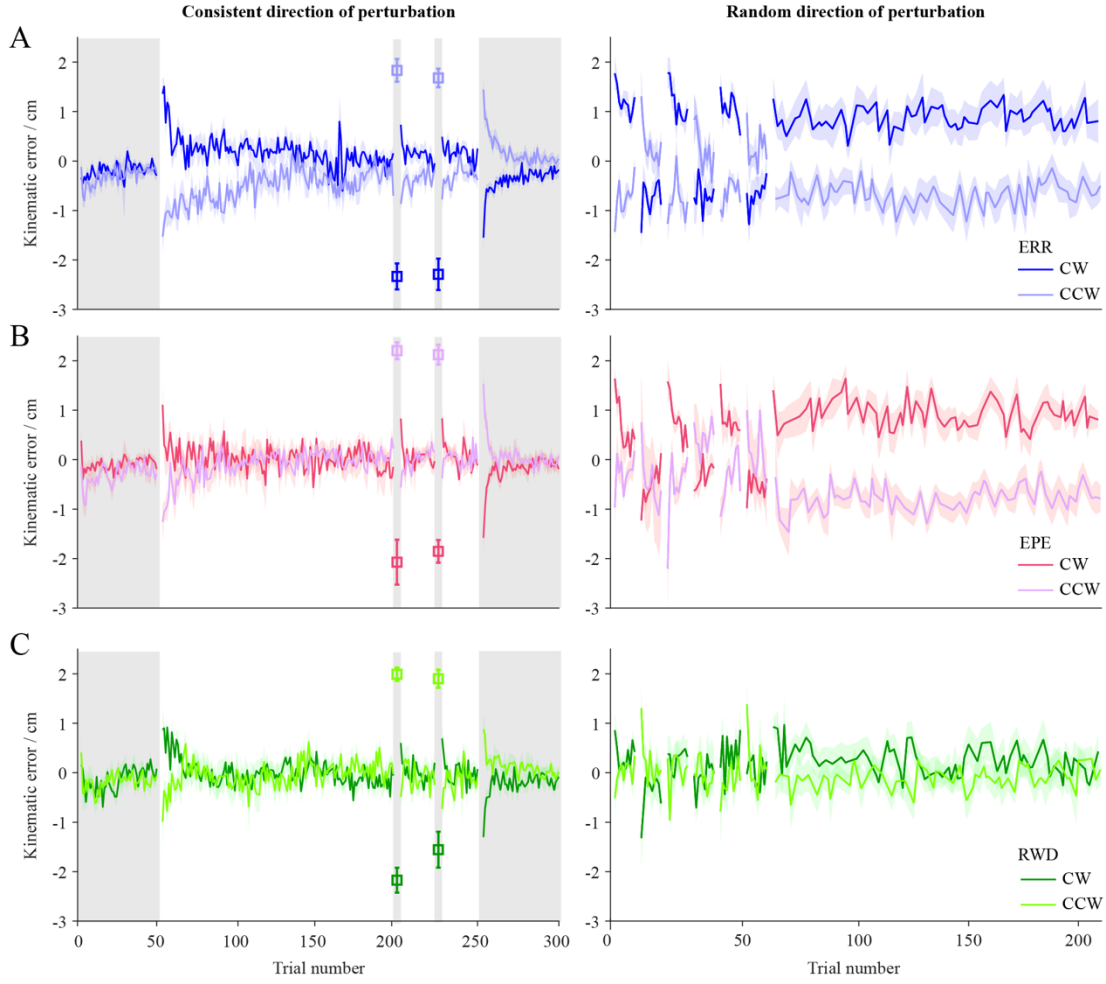


Figure 3.2: Kinematic error throughout the experiment with a single target location. Mean values of kinematic error for ERR group (A), EPE group (B) and RWD group (C). The left panels depict the initial part of the experiment where the direction of perturbation was consistent (both CW and CCW force fields for each participant). They show the increase in the kinematic error when perturbation is introduced, gradual learning with time and aftereffects during catch and de-adaptation trials. The right panels depict the increased kinematic errors during the final part of the experiment where the direction of perturbation was alternating. Dark colours indicate the kinematic error during the part when perturbation was in the clockwise direction whereas light colours indicate kinematic error during the part when perturbation was in the counter clockwise direction. Shaded area shows standard error of the mean. The grey shaded background indicates phases without perturbation whereas non-shaded background indicates the phases of the experiment with either clockwise or counterclockwise force-field perturbations.

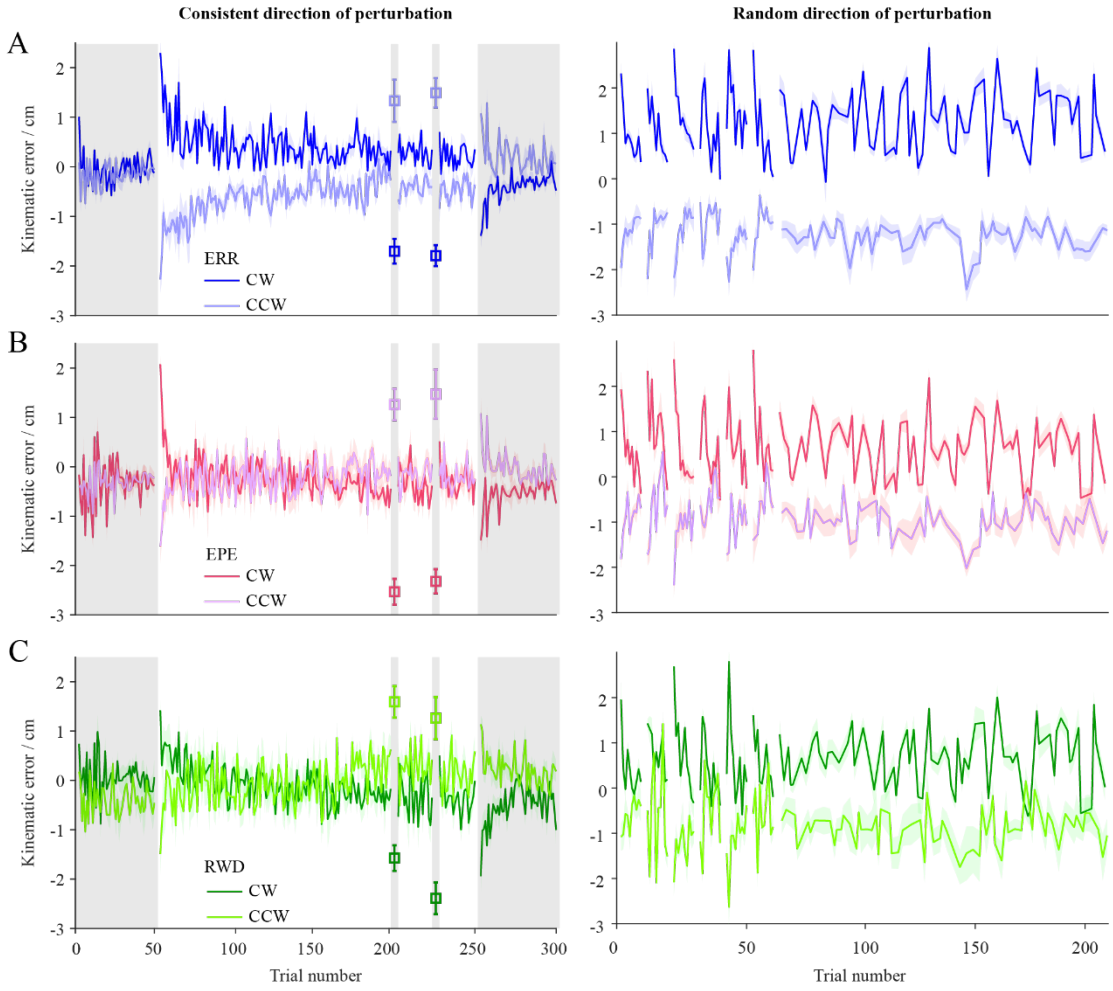


Figure 3.3: Kinematic error throughout the experiment with three target location. Mean values of kinematic error for ERR group (A), EPE group (B) and RWD group (C). The left panels depict the initial part of the experiment where the direction of perturbation was consistent (both CW and CCW force fields for each participant). They show the increase in the kinematic error when perturbation is introduced, gradual learning with time and aftereffects during catch and de-adaptation trials. The right panels depict the increased

kinematic errors during the final part of the experiment where the direction of perturbation was alternating. Dark colours indicate the kinematic error during the part when perturbation was in the clockwise direction whereas light colours indicate kinematic error during the part when perturbation was in the counter clockwise direction. Shaded area shows standard error of the mean. The grey shaded background indicates phases without perturbation whereas non-shaded background indicates the phases of the experiment with either clockwise or counterclockwise force-field perturbations.

3.2.1 Part 1: Consistent Direction of Perturbation

The objective of the initial phase of our study was to examine how participants adapted to consistent disturbances produced by a force field and to investigate potential differences in motor learning across three groups that were provided with different error information. Before each trial, all participants were provided with visual cues indicating the direction of the perturbation.

During the initial baseline movements, when there was no perturbation, the ERR, EPE and RWD groups exhibited comparable movements with nearly straight-line hand trajectories and the kinematic error values close to zero. All groups managed to predominantly hit the target area with the end-point error values inside the 0.7 cm target radius. When the perturbation was first introduced, the kinematic error increased in the direction of perturbation. After a few trials, kinematic errors decreased as participants proceeded to learn. Figure 3.4 shows the kinematic and end-point errors during the learning period. Participants in all groups adapted to force-field perturbation, which can be seen as a decrease in kinematic error during perturbation trials. In effect, hand trajectories became similar to the baseline hand trajectories with kinematic error coming close to zero for all groups.

We compared kinematic errors and end-point errors from last perturbed trials with baseline trials separately for each group to determine if participants fully adapted to perturbation. Movements in the ERR groups during last perturbation trials shifted in the same direction as perturbation indicating undercompensation. Even though kinematic errors during the experiment with single target location were comparable between baseline and late learning (baseline: (mean ± SD) 0.05 ± 0.11 cm ; late learning: 0.12 ± 0.24 cm; $t(9) = 0.75$, $p = 1$), the end-point errors during late learning were higher compared to baseline (baseline: 0.05 ± 0.07 cm; late learning: 0.21 ± 0.07 cm; $t(9) = 4.25$, $p = .013$). Kinematic and end-point errors during baseline were also higher compared to late learning in the experiment with 3 target locations (baseline: 0.07 ± 0.11 cm; late learning: 0.29 ± 0.15 cm; $t(9) = 5.09$, $p = .004$; baseline: 0.04 ± 0.05 cm, late learning: 0.20 ± 0.09 cm, $t(19) = 6.27$, $p < .001$). Movements in the EPE groups during last perturbation did not differ from movements during baseline. Kinematic and end-point errors during late learning trials were not statistically different from those during baseline for experiments with single target location (kinematic error: baseline: 0.08 ± 0.17 cm; late learning: -0.07 ± 0.29 cm; $t(9) = 1.28$, $p = 1$; end-point error: baseline: 0.15 ± 0.17 cm, late learning: 0.17 ± 0.23 cm, $t(9) = 0.21$, $p = 1$) and three target locations (kinematic error: baseline: -0.01 ± 0.13 cm; late learning: -0.19 ± 0.28 cm; $t(9) = 2.02$, $p = .447$; end-point error: baseline: -0.04 ±0.25 cm, late learning: 0.08 ± 0.16 cm, $t(9) = 1.14$, $p = 1$) for experiments with three

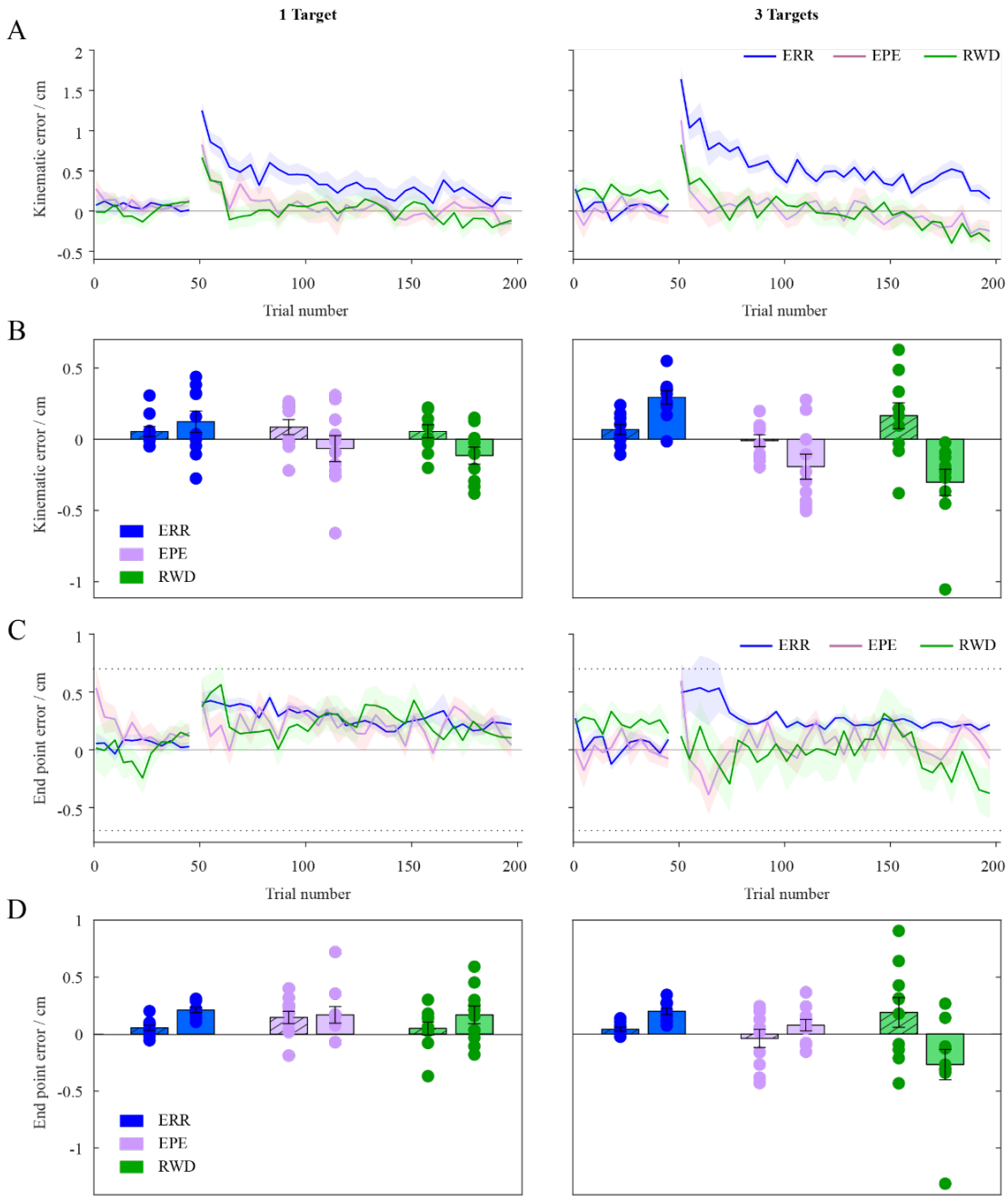


Figure 3.4: Adaptation to consistent force-field perturbation. Mean values of kinematic (A) and end-point (C) error showing the progression of learning combined for clockwise and counter-clockwise perturbation. All values are averaged across 4 trials. Shaded area shows standard error of mean. Grey dotted lines indicate the ±0.7 cm boundary corresponding to the target radius, indicating successful performance. (B, D) Comparison of kinematic and end-point errors between baseline and late learning phases within each group to assess adaptation to force-field perturbation, including indications of undercompensation and overcompensation. Kinematic and end-point error is averaged across all baseline trials (first bar, striped pattern) and last 20 perturbation trials (second bar, no pattern). Error bars show standard error of mean.

target locations. Moreover, movements in the RWD group in the experiment with single target location did not significantly differ between late learning and baseline period. Kinematic and end-point errors during late learning trials were not statistically different from those during baseline (Kinematic error: baseline: 0.06 ± 0.15 cm; late learning: -0.11 ± 0.19 cm, $t(9) = 2.23$, $p = .315$; end-point: baseline: 0.05 ± 0.18 cm, late learning: 0.17 ± 0.25 cm, $t(9) = 1.03$, $p = 1$). On the other hand, kinematic errors in the RWD group in the experiment with three target locations shifted slightly in the opposite direction as the perturbation, indicating possible overcompensation. Kinematic and end-point errors decreased from baseline to late learning, however after correcting for multiple comparisons, this difference was only statistically significant for the kinematic error (kinematic error: baseline: 0.17 ± 0.29 cm; late learning: -0.30 ± 0.30 cm; $t(9) = 3.46$, $p = .043$; end-point error: baseline: 0.19 ± 0.41 cm, late learning: -0.27 ± 0.42 cm, $t(9) = 2.60$, $p = .174$).

To further assess learning, we looked at the kinematic error of catch trials when the perturbation was absent. Kinematic error values of catch trials were comparable between all groups. There was no significant difference between groups for the first ($F(2) = 0.46$, $p = .633$) as well as the second catch trial ($F(2) = 0.77$, $p = .468$). Kinematic error values of catch trials in the experiment with 3 target locations were also comparable between all groups. There was no significant difference between groups for the first ($F(2) = 1.87$, $p = .164$) as well as the second catch trial ($F(2) = 0.07$, $p = .934$).

In addition to kinematic and end-point errors, compensatory force measured during channel trials provides a direct estimate of predictive, feedforward compensation to the force field. Normal channel trials reflect the overall expressed adaptation at that point in learning. In contrast, in instructed channel trials participants are cued to expect no perturbation, which minimizes explicit compensation and therefore serves as an index of implicit adaptation. We compared compensatory force measured during normal and instructed channel trials at the end of the learning period (Figure 3.5) across groups and number of targets. We found a significant interaction between group and type of channel trial ($F(2, 54) = 6.44$, $p = .003$). However, there was no interaction effect between group, number of targets and type of channel trial ($F(2, 54) = 1.09$, $p = .343$), no interaction effect between target and type of channel trial ($F(1, 54) = 3.00$, $p = .089$), no interaction between group and number of targets ($F(2, 54) = 1.83$, $p = .170$) and no main effect of target ($F(1, 54) = 1.20$, $p = .279$). This indicated that only group and type of channel trial and not the number of targets had an effect on the measured compensatory force. Therefore, we compared force compensation during channel trials across groups for the experiment with single and three target locations combined. For the instructed channel trials, there was no difference in compensatory force between the groups (ERR: 47 ± 14%, EPE: 57 ± 19%, RWD: 48 ± 26%; $F(2, 59) = 1.69$, $p = .193$) indicating the groups did not significantly differ in their levels of implicit adaptation. We also compared compensatory force measured during normal channel trials at the end of the learning period across groups and found significant differences ($F(2, 59) = 14.46$, $p < .001$). Average force compensation (ERR: 62 ± 12%, EPE: 82 ± 13%, RWD: 88 ± 20%) for ERR group was significantly lower compared to EPE ($t(38) = 4.80$, $p < .001$) and RWD ($t(38) = 4.82$, $p < .001$) groups but there was no difference in compensatory force between EPE and RWD groups ($t(38) = 1.11$, $p = .275$). This indicated higher contributions of explicit learning in EPE and RWD groups.

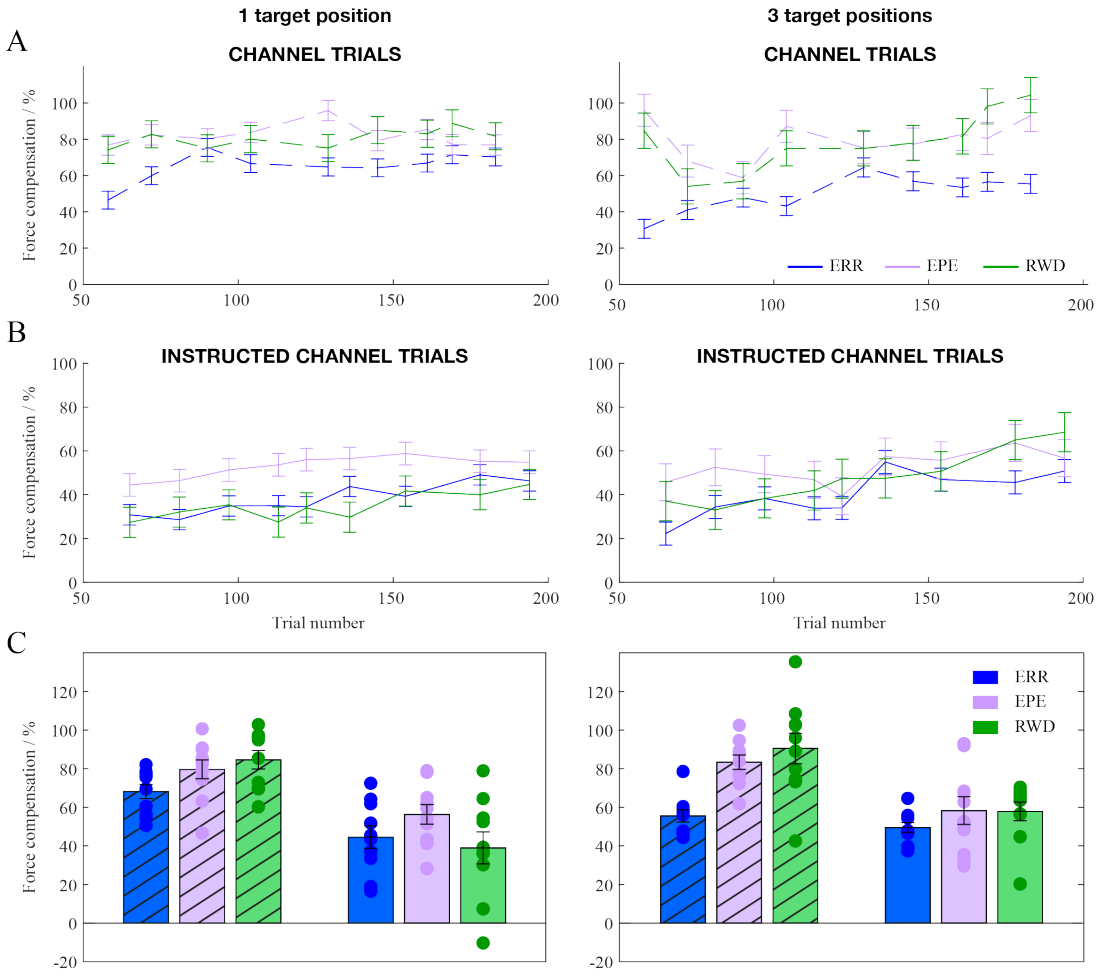


Figure 3.5: Force compensation during consistent direction of perturbation. (A) Mean values of force compensation measured during normal channel trials, showing differences in adaptation across groups. (B) Mean values of force compensation measured during instructed channel trials, showing the implicit component of adaptation. (C) Comparison of average force compensation across groups, indicating similar levels of implicit adaptation and differences in overall adaptation between groups. Force compensation was averaged over the last 4 trials. The striped pattern indicates normal channel trials and no pattern indicates instructed channel trials. Values are combined for both directions of perturbation showing the progression of learning separately for single target and 3 target locations. Error bars show standard error of mean.

Towards the end of learning, participants experienced two types of trials where perturbation was absent: catch and de-adaptation trials. During the catch trials, the force field perturbation was omitted without participants' knowledge (cue indicated the presence and direction of the force field). Contrarily, during the de-adaptation trials, participants were informed the perturbation would be absent (cue indicated that no force field was present). To assess the two types of after-effects, we compared the kinematic error of catch trials (trials 202 and 226, Figure 3.6) and 1st de-adaptation trials (trial 251, Figure 3.6) across different groups and number of targets. Statistical analysis showed only a main effect

of after-effect trial (catch vs de-adaptation trial) ($F(1, 54) = 43.71$, $p < .001$) indicating a smaller after-effect (smaller absolute kinematic error) during de-adaptation trials compared to catch trials. All other interaction and main effects were not statistically significant (interaction of group, number of targets and after-effect: $F(2, 54) = 2.93$, $p = .062$; interaction of number of targets and after-effect: $F(1, 54) = 3.70$, $p = .060$; interaction of group and after-effect: $F(2, 54) = 0.29$, $p = .749$; interaction of group and number of targets: $F(2, 54) = 0.21$, $p = .813$; main effect of number of targets: $F(1, 54) = 2.71$, $p = .105$; main effect of group: $F(2, 54) = 1.13$, $p = .332$). This indicates there was no statistical difference in kinematic error during catch trials between groups, nor a statistical difference in kinematic error during 1st de-adaptation trials between groups. However, it should be pointed out that the kinematic error during the 1st de-adaptation trial of RWD group (-1.09 ± 0.65 cm) during the experiment with single target location appeared smaller (in absolute terms) compared to EPE (-1.56 ± 0.60 cm) and ERR group (-1.50 ± 0.61 cm), though not statistically significant.

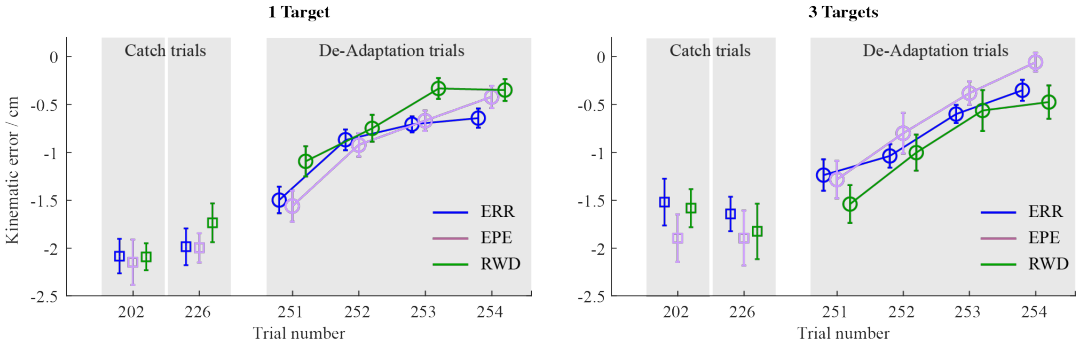


Figure 3.6: After-effects following adaptation to consistent direction of perturbation. Mean kinematic error for two catch trials and first four de-adaptation trials emphasizing the differences between the two types of after-effects. Catch trials involved unexpected removal of the perturbation, while de-adaptation trials were explicitly cued as trials without perturbation. Error bars show standard error of mean. Grey shaded blocks represent parts without perturbation.

3.2.2 Part 2: Alternating Direction of Perturbation

The objective of the second phase of our study was to examine how participants adapted to changing disturbances and investigate how different error information affected the ability to successfully perform the motion when the direction of perturbation switched randomly from trial to trial. Participants first performed 6 blocks of 10 trials with alternating clockwise and counter clockwise force field perturbations before proceeding to the random direction of perturbation phase. Figure 3.7 shows trials when force-field direction is alternating randomly. Trials were conducted in a random sequence; however, all participants were informed about the direction of the perturbation through visual cues before each trial. Our goal was to assess whether participants could effectively adjust their movements to counteract two opposing perturbations, relying on the advance information provided about the perturbation.

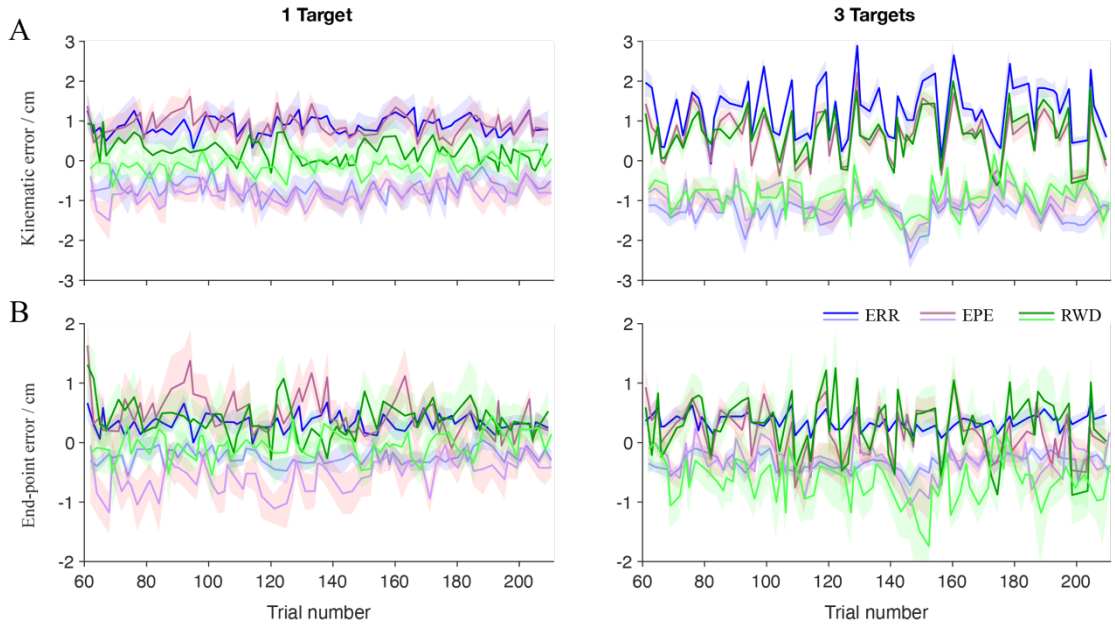


Figure 3.7: Adaptation to random force-field perturbation. Mean values of kinematic error (A) and end-point error (B) across participants for single target and 3 target locations showing the difference between movements with CW and CCW perturbation. Participants were informed about the direction of the perturbation through visual cues before each trial. Shaded area shows standard error of mean. Measures in the CW force field (dark lines) and the CCW force field (light lines) are shown separately.

We compared the kinematic errors between clockwise and counter clockwise perturbation for each group. The ERR and EPE groups had significantly different kinematic errors between the two directions of perturbation (ERR: CCW: -0.67 ± 0.65 cm, CW: 0.87 ± 0.87 cm, $t(9) = 3.34$, $p = .017$; EPE: CCW: -0.80 ± 0.47 cm, CW: 0.88 ± 0.52 cm, $t(9) = 7.04$, $p < .001$) in the experiment with single target location. Moreover, in the experiment with three target locations, almost all groups had significantly different kinematic errors between the two directions of perturbation (ERR: CCW: -1.26 ± 0.49 cm, CW: 1.26 ± 0.36 cm, $t(9) = 10.55$, $p < .001$; EPE: CCW: -0.99 ± 0.52 cm, CW: 0.61 ± 0.22 cm, $t(9) = 7.64$, $p < .001$; RWD: CCW: -0.89 ± 0.40 cm, CW: 0.65 ± 0.27 cm, $t(9) = 8.64$, $p < .001$). Interestingly, the kinematic errors for RWD group did not significantly differ between the two force-field directions (CCW: -0.10 ± 0.58 cm, CW: 0.24 ± 0.64 cm, $t(9) = 0.97$, $p = .356$).

Additionally, we tested the effect of group and number of targets on the average kinematic error during the random direction of perturbation. To do this, we appropriately flipped the sign of the errors in the CCW force field such that a positive value indicated an error produced by the force field, prior to averaging with the CW force field data. There was no significant interaction effect ($F(2, 59) = 2.60$, $p = .084$), however there was a significant main effect of group ($F(2, 59) = 8.59$, $p = .005$) and significant main effect of number of targets ($F(1, 59) = 6.99$, $p = .002$). The corresponding post hoc tests showed that the average kinematic errors in the experiment with single target location were significantly smaller than the average kinematics errors in the experiment with three target

locations (single target location: 0.59 ± 0.63 cm, three target locations: 0.95 ± 0.39 cm, $t(58) = 2.61$, $p = .035$). Moreover, kinematic errors of the RWD group (0.47 ± 0.53 cm) were significantly smaller compared to the EPE (0.82 ± 0.34 cm) ($t(58) = 2.47$, $p = .036$) and ERR (1.02 ± 0.62 cm) ($t(58) = 3.00$, $p = .019$) groups across experiments with single target and three target locations.

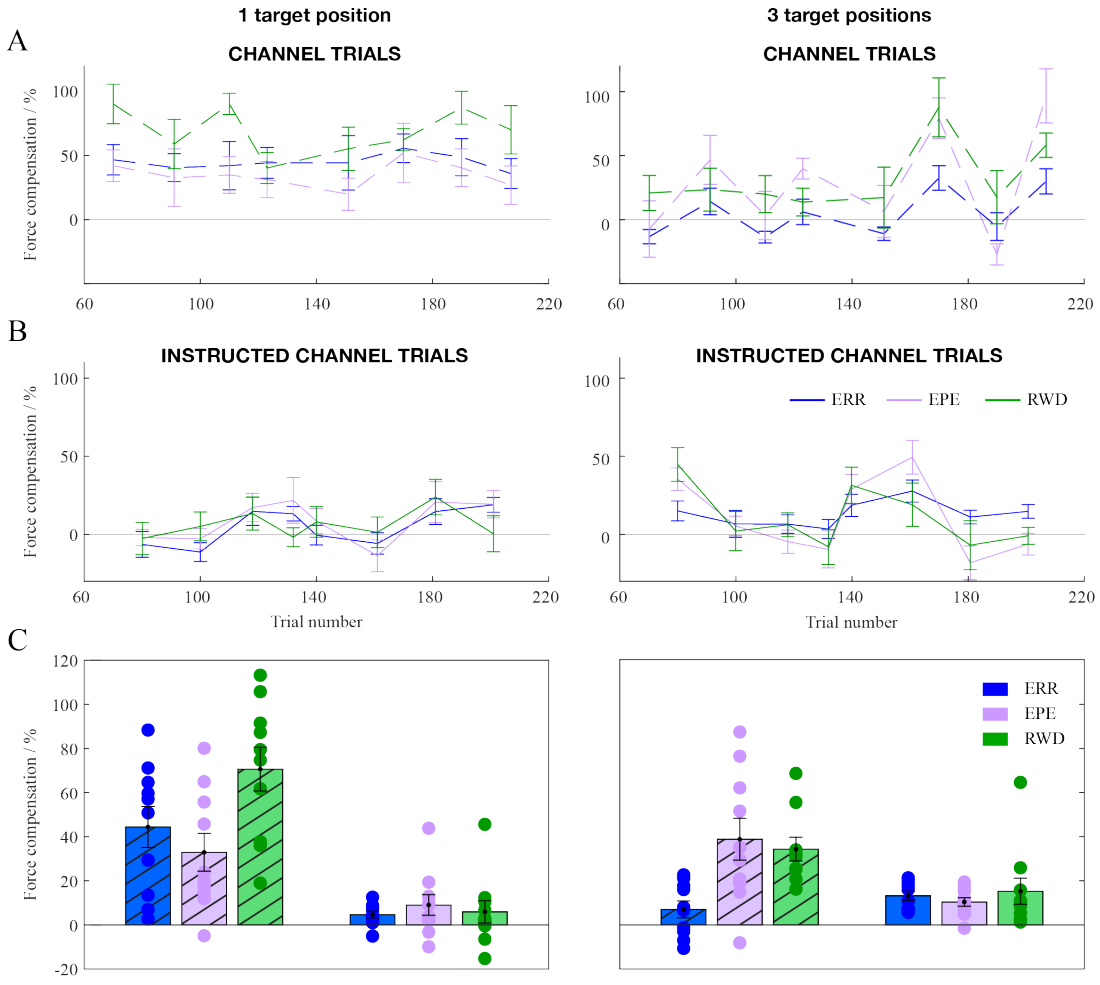


Figure 3.8: Force compensation during random direction of perturbation. (A) Mean values of force compensation measured during normal channel trials, showing differences in adaptation across groups. (B) Mean values of force compensation measured during instructed channel trials, showing the implicit component of adaptation. (C) Comparison of average force compensation across groups, indicating similar low levels of implicit adaptation and differences in overall adaptation between groups. Force compensation was averaged over all trials. The striped pattern indicates normal channel trials and no pattern indicates instructed channel trials. Values are combined for both directions of perturbation showing adaptation separately for single target and 3 target locations. Error bars show standard error of mean.

We compared the force compensation measured during normal and instructed channel trials across groups and number of targets (Figure 3.8) finding a significant interaction

effect ($F(2, 53) = 6.31$, $p = .003$). We then compared compensatory forces between groups separately for both single and 3 target locations and separately for normal and instructed channel trials. During instructed channel trials, there was no difference in compensatory force between groups for both experiments with single (ERR: 5 ± 5%, EPE: 9 ± 15%, RWD: 6 ± 16%; $F(2, 29) = 0.30$, $p = .743$) and three (ERR: 13 ± 6%, EPE: 10 ± 6%, RWD: 15 ± 19%; $F(2, 29) = 0.41$, $p = .665$) target locations suggesting groups did not significantly differ in their levels of implicit adaptation. Overall, the force compensation on the instructed channel trials was close to zero across all groups and for both one target and three target experiments. On the other hand, there were significant differences between groups during normal channel trials in the experiment with single target location (ERR: 44 ± 29%, EPE: 33 ± 27%, RWD: 70 ± 31%; $F(2, 29) = 4.32$, $p = .024$) and three target locations (ERR: 7 ± 12%, EPE: 39 ± 30%, RWD: 34 ± 17%; $F(2, 28) = 6.51$, $p = .005$). During normal channel trials, the RWD group had higher compensatory forces compared to EPE group ($t(18) = 2.87$, $p = .041$) in the experiment with single target location. Moreover, both RWD ($t(17) = 4.07$, $p = .005$) and EPE ($t(18) = 3.11$, $p = .030$) groups had higher compensatory forces compared to ERR group in the experiment with three target locations.

3.3 Discussion

To compare and analyse how learning is affected by the presentation of different error signals, we employed a force-field perturbation paradigm with contextual switching. Participants performed arm reaching movements under two different force field perturbations while they were presented either with the sensory prediction error or task error (task error could be in the form of end-point error or binary error). In all cases, participants were provided with a visual signal indicating the presence and direction of the force field on the subsequent trial. This approach allowed us to assess how each type of error feedback influenced motor adaptation. Our results showed that learning based on sensory prediction error leads to undercompensation, while learning based on task error induces complete compensation or even overcompensation due to the higher contribution of explicit strategies. Moreover, the results of the contextual switching task showed that only participants learning based on binary information exhibited some degree of adaptation to random perturbations, relying almost entirely on explicit strategies.

In the group that received full visual feedback throughout the movement and learned based on the sensory prediction error, there were small kinematic and end-point errors present at the end of the consistent learning period. These errors were in the same direction as the perturbation, indicating undercompensation. The errors were not present in the two groups that learned based on the task error alone. The full visual feedback group also had smaller force compensation compared to the other two groups that learned based on the task error. Incomplete compensation was previously shown in motor adaptation, where small residual errors persisted despite practice (Vaswani et al., 2015; Weightman et al., 2022). Residual errors are a feature of implicit learning and arise because of interaction of learning and forgetting (Albert et al., 2021; Smith et al., 2006). They are not present when learning relies on explicit strategies (Albert et al., 2021; Langsdorf et al., 2021). For the

most part, the group that learned based on the task error in the form of end-point error, reached accurate compensation, where movements at the end of the learning period did not significantly differ from the baseline movements.

Some overcompensation was observed in the group that learned based on the task error in the form of binary information. In the 3 target condition, kinematic and end-point errors at the end of the learning period appeared to shift in the opposite direction of the perturbation, suggesting a possible overcompensation effect. This overcompensation is only observed in kinematic data and not evident in force compensation, since force measures may not fully reflect all learning processes, including feedback responses. Overcompensation may indicate interaction of implicit and explicit learning; gradual implicit adaptation that operates independently and in parallel to the explicit strategy (Albert et al., 2022; Mazzoni & Krakauer, 2006b; Taylor & Ivry, 2011). Usually, we can see the amount of error in our movements and adjust our explicit strategies accordingly. However, in this case participants only received binary information and were unaware of their exact position relative to the centre of the target. As a result, they could not adjust their strategies effectively and their movements may have gradually shifted toward the target perimeter. They only realized they were making increasing errors when they received negative feedback after moving outside the target.

Interestingly, we found that compensatory force during the instructed channel trials did not statistically differ between groups, indicating that levels of implicit learning are not significantly different across different types of error information provided. However, groups that learned based on task information alone exhibited a significantly higher compensatory force during normal channel trials. Previous studies have often modelled adaptation as the sum of implicit and explicit learning (Miyamoto et al., 2020), though we recognize this is a simplification and does not fully capture all contributing factors. Given the comparable levels of implicit learning, our findings suggest that learning based on task error engages additional explicit strategies compared to learning based on sensory prediction error. This may indicate that the amount of explicit learning is what primarily drives the difference in adaptation based on the specific type of error information. However, other mechanisms, such as feedback-driven responses, may also contribute to differences in adaptation across conditions.

Participants that learned based on sensory prediction error exhibited larger kinematic errors during random direction of perturbation compared to kinematic errors during consistent direction of perturbation suggesting sensorimotor learning alone does not enable simultaneous adaptation to two opposing perturbations, which is in line with previous research (Howard et al., 2013; Karniel & Mussa-Ivaldi, 2002; Kunavar et al., 2023). However, despite the large kinematic errors during random direction of perturbation, participants made accurate corrections during the movement and still managed to successfully hit the target as reflected in the end-point errors which were inside the target radius. The fact that they were able to consistently hit the target without reducing the kinematic error could indicate that relying on feedback might have been a more optimal solution for them. However, since the adaptation to the consistent direction of perturbation showed relatively straight trajectories with small kinematic errors, we could argue that the large kinematic errors are a consequence of the inability to switch between motor

commands obtained for two different movements rather than feedback alone being a more optimal solution. Another possibility would be that switching between different motor plans or strategies is too costly compared to feedback corrections, which remain more feasible when full visual feedback is available. Evidence from motor learning research suggests that switching between motor plans can introduce measurable performance costs (Kiesel et al., 2010; Orban de Xivry & Lefèvre, 2016). Additionally, cognitive load may play a role, as explicit strategy use during motor adaptation relies on working memory, which has limited capacity (Christou et al., 2016).

On the other hand, participants that learned based on task error in the form of binary information, which relies more on the use of explicit strategies (Spampinato & Celnik, 2021; Taylor et al., 2014), were better at adapting to the random direction of perturbation especially during the experiment with a single target location. This is in line with studies that showed that dual adaptation was possible primarily through explicit strategies (Hegele & Heuer, 2010; Kunavar et al., 2023; Schween et al., 2018). It is also possible that the different cost structures of errors in different feedback conditions influenced how participants adapted. In the continuous feedback condition, where full information allows for gradual corrections, small errors may have been more tolerable. In contrast, in the binary feedback condition, the cost of missing the target was the same for small and large errors, potentially encouraging a stronger reliance on explicit strategies to minimize target misses. This suggests that adaptation differences between conditions may be due to differences in control strategies driven by error cost considerations. It should be pointed out that the adaptation to opposing perturbations followed a learning period with consistent direction of perturbation, therefore participants were able to use previously learned motor behaviour. It should be further tested if the differences between learning based on different error information would translate to direct adaptation to opposing direction of perturbation.

Similar as in previous studies (Forano et al., 2021; Taylor et al., 2014), the ability to implement explicit strategies decreased when multiple targets were introduced. This is likely due to the restricted capacity of conscious learning processes and limited working memory that does not allow for extracting and storing high amount of information (Forano et al., 2021). Although the targets in our experiment were only 45 degrees apart, making explicit strategy use still feasible to some extent, our results indicate that adaptation was noticeably more difficult in this condition compared to the single-target case. A crucial factor to consider is that participants had to adapt to two perturbations, resulting in six different strategies that needed to be managed.

We would expect that learning based on task error in the form of end-point error generated similar results as learning based on task error in the form of binary information, however participants that learned based on end-point error did not manage to adapt to the random direction of perturbation and exhibited similar kinematic errors as participants learning based on sensory prediction error. This is contrary to the results obtained by a visuomotor study where groups that learned based on end-point error and binary information exhibited similar results and both managed to adapt to the random direction of perturbation (Kunavar et al., 2023). However, it is important to note that our study did not directly compare these conditions across paradigms, and that both experiments only

measured movement errors, not forces, which limits direct comparisons of underlying learning processes. The discrepancy between visuomotor rotation and force-field perturbation results might be due to the adaptation being driven by either visual or proprioceptive errors. One possible explanation is that, in the case of force-field experiments, adaptation may be primarily driven by proprioceptive errors, and the combination of proprioceptive information with visual end-point feedback may not necessitate a greater degree of explicit learning. In contrast, in the case of binary feedback, the lack of specific error information may increase reliance on explicit strategies. Additionally, the type of adaptation needed in force-field adaptation compared to visuomotor adaptation differs. Visuomotor rotation requires just a single change in reaching direction for adaptation and therefore a fast adaptation can give perfect results. On the other hand, adaptation to a force field requires a complex pattern of adaptation throughout the movement and makes it much harder to rely solely on explicit adaptation. However, the exact mechanisms behind these differences are not well understood, and this remains a speculative explanation. Further research is needed to investigate how proprioceptive and visual feedback interact in motor learning and adaptation.

In conclusion, our study reveals how different types of error feedback—sensory prediction, end-point, and binary task error—affect motor adaptation in response to force-field perturbations. We found that all groups displayed similar levels of implicit learning, while task error feedback enhanced the use of explicit strategy. This led to overcompensation due to greater reliance on explicit strategies or undercompensation due to combination of forgetting and smaller reliance on explicit strategies. Explicit strategies, enhanced by binary feedback, enabled better switching between opposing perturbations, though cognitive load limited their use with multiple target locations. The study also suggests that task specificity, particularly the difference between visual and proprioceptive error processing, influences the balance between implicit and explicit learning.

3.4 Methods

3.4.1 Participants

Sixty right-handed participants participated in the study. They were divided into three groups, the ERR group (n = 20, 8 female, 12 male, age 22.35 ± 2.20 years), the EPE group (n = 20, 9 female, 11 male, age 22.40 ± 2.52 years) and the RWD group (n = 20, 9 female, 11 male, age 22.45 ± 3.06 years). Each group was further divided into two subgroups, where the first subgroup always reached to a single target location and the second subgroup reached to 3 different target locations during the experiment. Prior to the experiment, the participants were informed about the course of the experiment and signed a written consent. All experimental protocols were approved by the National Medical Ethics Committee Slovenia (0120-348/2023/6) and the methods were carried out in accordance with the relevant guidelines.

3.4.2 Experiment

Participants were seated in a chair in front of a table with an integrated screen. They held the handle of a haptic robot (Haptic Master MK2, Moog, Nieuw-Vennep, The Netherlands) that was located underneath the table and not visible to the participants. The participants were asked to perform a series of arm reaching movements. The experiment started with the Haptic Master positioning the handle in a start position. The start position was displayed on a screen as a white circle with a diameter of 0.8 cm. The position of the participant's hand was displayed on the screen as a cursor in a form of a red dot. The target towards which the participants had to perform the arm movement was a blue circle with a diameter of 1.4 cm which was located 10 cm away from the start position. The number of possible target locations was either 1 or 3. The single target was located directly in front of the start position, whereas the 3 target group also included two targets angled at ±45°. The ERR group was provided with sensory prediction errors by showing them the cursor during the whole reaching task. Here we assumed that the participants primarily rely on the sensory prediction error since in the presence of sensory information participants predominantly disregard the information of the task success (Izawa & Shadmehr, 2011). Distinct visual information presented to different groups is shown on Figure 3.1A. Both the EPE and RWD groups had the cursor hidden during the reaching motion. The EPE group was provided with the information on where they passed the distance of the target (end-point error) and whether the target was hit or missed. The RWD group only received information whether the target was hit or missed. Since there was no visual feedback during the movement, information on sensory prediction error was minimized and participants should only rely on information regarding the task success.

There were three types of trials: unperturbed, perturbed and channel trials Figure 3.1B. Two opposing velocity dependent curl force fields were used as a perturbation. The force at the handle was calculated as:

$$\begin{bmatrix} F_x \\ F_y \end{bmatrix} = \begin{bmatrix} 0 & k \\ -k & 0 \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} \tag{3.1}$$

where F_x and F_y are forces applied to the hand, $\dot{x}$ and $\dot{y}$ are hand velocities and k is 15 or -15 Ns/m for force field in clockwise or counter clockwise direction (Gandolfo et al., 1996). During the channel trials, the reaching movement was confined to an emulated mechanical channel with a stiffness of 5,000 N/m acting lateral to the line from the start position to the target (Milner & Franklin, 2005; Scheidt et al., 2000).

Prior to each trial a visual cue was displayed at the centre of the screen to indicate the presence and direction of the perturbation; an x sign for unperturbed trials, right arrow for perturbations in a clockwise direction, and left arrow for perturbations in a counter clockwise direction. To enhance the visibility and distinction of these cues, we utilized varying background colours for different experimental conditions. The purpose of these visual cues was to facilitate the measurement of feedforward learning, as opposed to reactive feedback responses. Visual cues were consistently provided across all trials and on most trials, they correctly represented the actual perturbation. There were two cases where the visual cues were not correct; catch trials and channel trials. During catch trials, the

perturbation was absent even though participants still received visual cue indicating perturbation. Similarly, during channel trials, participants received visual cue indicating perturbation while the reaching movement was confined to a simulated mechanical channel. During instructed channel trials, the reaching movement was confined to a simulated mechanical channel and participants were informed that the perturbation was removed. Participants were briefed on the occurrence of both perturbed and unperturbed trials. They were informed that visual cues signalling the presence or absence of perturbation would precede each trial. The discontinuation of perturbation was also communicated through visual cues, ensuring participants were well-informed about the experimental conditions.

3.4.3 Experimental Protocol

The experiment was divided into two consecutive parts as shown on Figure 3.1C. In the first part, the direction of perturbation was consistent, while in the second part, the direction of perturbation was alternating between clockwise and counter clockwise. Two 40s pauses occurred between blocks, and a 10s pause was given every 60 trials.

3.4.3.1 Consistent Direction of Perturbation

The first part with the consistent direction of force field perturbation was intended for participants to adapt to a new perturbation. Half of the participants started with the clockwise perturbation and half with the counter clockwise perturbation. Participants first performed 50 baseline trials without perturbation, followed by 200 trials with a consistent perturbation in a single direction and 50 de-adaptation trials without perturbation. The experiment was then repeated with the other direction of perturbation, where participants performed 30 baseline trials, 200 trials with perturbation and 50 trials without perturbation. To assess the adaptation to perturbation, we included 2 catch trials towards the end of the perturbed trials. Channel trials and instructed channel trials were present approximately every 10 trials. Overall, we had 37 normal catch trials and 22 instructed channel trials across both directions of perturbation. During the baseline trials, participants had 700 ± 100 ms to perform the movement, otherwise they received a written message “too fast” or “too slow”. In other trials the time limit was extended to 700 ± 200 ms, to avoid participants being too focused on the time limit.

3.4.3.2 Alternating Direction of Perturbation

The second part of the experiment was designed to measure adaptation to the alternating direction of the perturbation. It started with 6 blocks of 10 trials where direction of perturbation changed from one block to another. The direction of perturbation was counterbalanced across participants. Following, the participants had to perform a randomly ordered set of 150 trials with either the clockwise or counter clockwise perturbation. Participants were informed about the direction of the perturbation through visual cues before each trial. The pattern of the perturbation directions was consistent for all participants during this set. We also included 8 normal channel trials and 8 instructed channel trials equally spaced across the set of trials with random alterations of perturbation.

3.4.4 Data Analysis

Position, velocity and forces of the hand during the arm reaching motion were recorded by the haptic robot with a frequency of 100 Hz. Motions were segmented into separate trials where the start of the trial was defined as the moment when the hand position left the starting position and the trial concluded when the hand position reached a distance of 10 cm relative to the starting point. To quantify the effects of perturbation on the motion and the adaptation to perturbation over the trials, we calculated kinematic error and end-point error. Kinematic error was defined as the maximum perpendicular distance between the hand trajectory and a straight line between the start position and the target inside the 4 to 6 cm distance from the start target (the middle of the movement). End-point error was defined as a distance between hand trajectory and a target when the hand position reached the distance of 10 cm from the start position. To determine the amount of predictive force adaptation we calculated force compensation for every channel trial. Force compensation was calculated using a regression between the measured lateral force and the force needed to perfectly compensate for the perturbation (Smith et al., 2006). We removed outliers in force compensation data, where outliers were defined as 1.5 interquartile ranges above the upper quartile or below the lower quartile. With this, we removed 7% of force compensation data during normal channel trials and 2.4% of data during instructed channel trials.

Data were analysed offline using Matlab (R2019b, The MathWorks, Natick, MA). After each Analysis of Variance (ANOVA), further differences were examined using corresponding simple main effects and post-hoc comparisons with Holm-Bonferroni correction adjusting for multiple comparisons. We reported the adjusted p-values. Average values are reported in the format of mean ± standard deviation. For most of statistical analysis, kinematic errors and end-point errors during the part when perturbation was in the counter-clockwise direction were multiplied with -1, so that the results from both directions of perturbation were comparable. We averaged clockwise and counter-clockwise data during consistent direction of perturbation for each subject. To determine if participants fully adapted to perturbation, we compared kinematic errors and end-point errors from last perturbed trials (last 20 trials before catch trials) with all baseline trials for each group using paired t-tests. To assess the two types of after-effects, we compared the kinematic error of the two catch trials and 1st de-adaptation trial using repeated measures ANOVA with a main effect of after-effect (catch and de-adaptation trials) and between subject factors group (ERR, EPE and RWD) and number of targets (single and three target locations). ANOVA with a factor of group and number of targets was used to compare average kinematic errors during the random direction of perturbation. Additionally, we compared the kinematic errors between clockwise and counter clockwise perturbation during random direction of perturbation for each group using paired t-tests. Here, the counter-clockwise direction was not multiplied with -1. Force compensations at the end of the learning period (last 4 channel trials before catch trials) were compared with repeated measures ANOVA with a main effect of channel type (normal and instructed channel) and between subject factors group and number of targets. The same comparison was also used for force compensation during random direction of perturbation.

Chapter 4

Success-Efficient /Failure-Safe Strategy for Hierarchical Reinforcement Motor Learning

4.1 Introduction

Evolution has endowed us with motor learning abilities to boost our survival odds by enhancing competitive success (Krakauer et al., 2017). Ecological fitness refers to the specific traits of an organism and abilities that aid survival in its environment (Peacock, 2011). The sensorimotor system is crucial for ecological fitness, enabling adaptive movement control vital for survival (Faisal et al., 2008). Therefore, understanding sensorimotor adaptation and motor control optimality must consider ecological fitness. Overlooking this aspect leads to incomplete sensorimotor models that do not fully capture the brain-body-environment interaction. The adaptive capacity of the sensorimotor system, in addition to the evolutionary optimization that has occurred, is subject to the continued refinement of movement policy throughout an organism's life (Shadmehr & Krakauer, 2008). Thus, a thorough grasp of the sensorimotor system demands a multidisciplinary approach that embraces ecological fitness and integrates recent neuroscience and evolutionary biology insights (Wolpert et al., 2011).

Recent advancements in behavioural research have elucidated several distinct learning processes inherent in motor skill acquisition, involving a variety of neuronal substrates and computational processes (Haith & Krakauer, 2013a; Krakauer & Mazzoni, 2011; Shadmehr & Krakauer, 2008; Spampinato & Celnik, 2021; Taylor & Ivry, 2014; Wolpert et al., 1995). Among these, the basal ganglia are crucial for reinforcement learning (Doya, 1999; Kawato & Cortese, 2021; Schultz et al., 1997), which facilitates the refinement and acquisition of novel motor skills by providing feedback that rewards successful movements and discourages ineffective ones. Motor behaviour can thus be shaped through experience and practice. This is carried out in interaction with internal models of motor control, by updating predictions and adjustments based on feedback to enhance the ability of the brain to anticipate and execute precise motor actions (Kawato, 1999; Miall & Wolpert, 1996).

Internal models are neural representations that predict the sensory consequences of movements and generate feed-forward motor commands, enabling precise and efficient

motor control (Kawato, 1999; Kawato et al., 1987; Miall & Wolpert, 1996; Wolpert et al., 1998). These models are believed to reside partly in the cerebellum and are essential for human motor control. Past research has emphasized the role of the cerebellum in updating motor predictions based on sensory feedback to refine movement execution (Shadmehr & Krakauer, 2008; Wolpert et al., 2011). The brain also adjusts feedback control gains in response to movement errors, dynamically balancing predictive control with real-time corrections (Codol et al., 2023; Osu et al., 2002). Optimal feedback control theory explains how the brain computes motor commands in real time by balancing movement accuracy against energy costs (Scott, 2004; Todorov & Jordan, 2002), and empirical studies confirm that the human sensorimotor system operates according to these principles (Krakauer et al., 2017; Wolpert et al., 2011).

Beyond optimizing movement efficiency, motor control must integrate environmental risk assessment when planning and executing movements (Braun et al., 2011). Research in neuroeconomics suggests that the nervous system incorporates uncertainty into decision-making, dynamically adjusting motor strategies to balance performance with safety (Trommershäuser et al., 2009). This framework provides insights into the neural and computational mechanisms underlying motor planning in uncertain environments (Wolpert & Landy, 2012). Risk-sensitive motor adaptation is particularly evident in gait and postural control, where stability maintenance is crucial to prevent falls. Studies have shown that humans adjust step width, base of support, and foot placement in response to perceived instability (Finley et al., 2013). Similarly, research has demonstrated that metabolic cost interacts with risk perception in shaping movement strategies, emphasizing the continuous trade-off between efficiency and stability (Summerside et al., 2018).

We propose that the optimality concept in computational motor control needs a redefinition from an evolutionary perspective, where ecological fitness is central in guiding efficient movement generation. This approach would allow traditional cost functions, which relate to efficient movement execution, to be balanced alongside injury risk reduction. By movement efficiency we refer to the optimal use of muscles to perform movements with maximal performance and minimal energy expenditure. We hypothesize that the human brain employs a success-efficient/failure-safe strategy, wherein success reinforces energy-efficient movement execution, whereas failure enhances risk aversion and prioritizes safety. This approach suggests a hierarchical reinforcement learning mechanism that dynamically adjusts motor learning objectives based on past experiences.

To test our hypotheses, we developed a unique experimental approach that overcomes the limitations of prior studies by incorporating ecological fitness, specifically exposing the sensorimotor system to potential failure and injury risk. In our study, participants engaged in unrestrained whole-body motions, performing squat-to-stand movements under significant force perturbations. Unlike traditional paradigms that focus on isolated limb movements, this experiment examines whole-body motion, highlighting the ecological constraints that shape motor learning while introducing real-world risks such as falling and injury, which directly affect ecological fitness of participants. Moreover, the high energy demands of squat-to-stand movements necessitate a sophisticated neural strategy that balances energy optimization with stability maintenance, further emphasizing the brain's adaptive control mechanisms. By adopting this ecologically valid framework for studying

sensorimotor control, we seek to integrate human motor control into a broader hierarchical reinforcement learning model, where ecological fitness influences adaptation and learning throughout the life of an organism.

4.2 Results

A group of participants performed a series of whole-body stand-up motions (trials) while their centre of mass (COM) was systematically perturbed in the backward direction. On average the movement time was (Mean ± SEM) 1.16 ± 0.067 s. Perturbations were accomplished by a force-controlled pulling mechanism (Figure 4.1A) with a magnitude proportional to the vertical velocity of the COM. The protocol involved four types of trials; baseline, perturbed, catch, and de-adaptation trials (Figure 4.1B). The experiment began with a baseline block of 15 unperturbed trials, followed by a perturbation block consisting of a series of perturbed trials during which the participants were required to perform 60 successful trials. This was followed by a single catch trial, where the perturbation was turned off without the participants' awareness, and then by an additional five perturbed trials. The experiment concluded with a de-adaptation block of 15 unperturbed trials, where the participants were made aware that the perturbation had been turned off. To show the computational plausibility of our hypothesis, we developed a computational model and simulated the same experimental protocol as in the human experiments. The computational model has feedback gain tuning; internal model learning and movement policy learning integrated in a hierarchical learning system under ecological control. The model was tested with a simplified dynamic model of the human body.

For each trial, we obtained a COM trajectory based on the kinematic data and labelled it as a success or a failure (Figure 4.1C). A trial was considered a success if the participant successfully performed the stand-up motion. Conversely, the trial was considered a failure if the participant made a corrective step to avoid falling over. While corrective steps are classified as failures within the constraints of our study, it should be noted that they can be a key strategy for maintaining stability in natural movements (Bauby & Kuo, 2000; O'Connor & Kuo, 2009; Reimann et al., 2017). The overall effect of motor learning was assessed by calculating the trajectory area of the COM across the entire stand-up motion (Trajectory Area). The effect of motor planning was assessed by calculating trajectory area in the initial phase of the motion, well before the feedback mechanisms could alter the motion (Initial Trajectory Area). The Trajectory Area and Initial Trajectory Area were defined as the integral of the deviation of the movement trajectory from a straight line in the sagittal plane. The reference line was drawn from the initial COM height in the squat position to the final COM height in the standing position for Trajectory Area. Initial Trajectory Area was calculated during the first 5% of the motion, which occurred on average within (Mean ± SD) 96.4 ± 7.0 ms and corresponded to a height change of approximately 2.5 cm. The trial-by-trial progression of the average Trajectory Area and Initial Trajectory Area, for both successful and failed trials during the whole experiment, is shown in Figure 4.1D and E. Moreover, to determine the relation between the feed-forward and feedback aspects of motor control, we calculated the Smoothness of the COM trajectory, and the Co-Contraction of the thigh antagonistic pair of muscles. Specifically,

we used Smoothness to assess the accuracy of the internal model (Balasubramanian et al., 2015; Harris & Wolpert, 1998; Rohrer et al., 2002) and Co-Contraction to assess the contribution of feedback on motor control (Heald et al., 2018; Osu et al., 2002). It should be acknowledged that the level of muscle co-contraction has an effect on the jerk of the COM trajectory, and thus also influences Smoothness.

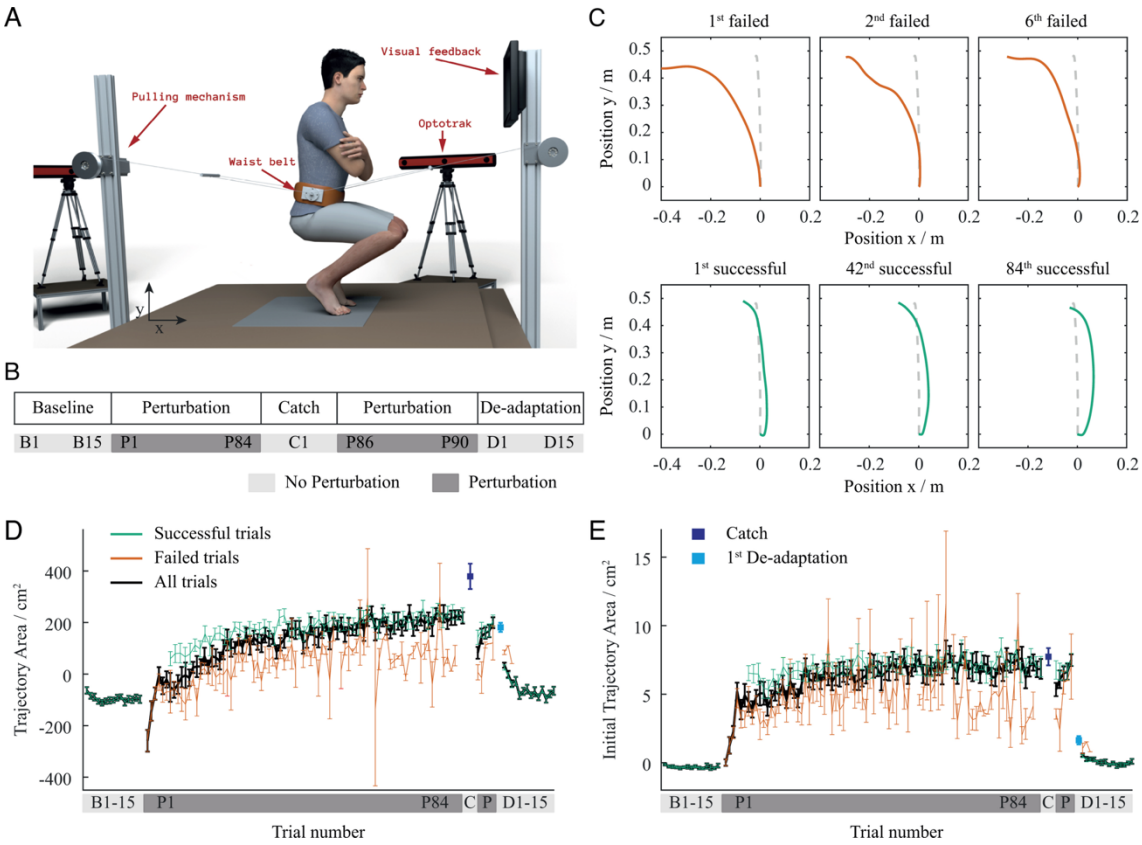


Figure 4.1: Experimental setup and centre of mass adaptation dynamics. (A) Illustration of the experimental setup where participants perform squat-to-stand movements using a custom force-controlled pulling mechanism to perturb their centre of mass. (B) Overview of the experimental protocol where light grey indicates unperturbed trials while dark grey indicates perturbed trials. (C) Failed and successful trajectories of a representative participant where dashed lines show baseline trajectory, red represents failed trials and grey represents successful trials. Trial-by-trial progression of Trajectory Area (D) and Initial Trajectory Area (E) during the whole experiment where red represents failed trials, green represents successful trials, black represents all trials, dark blue represents the catch trial, and light blue represents the first de-adaptation trial. Error bars show standard error of the mean.

Participants began the experiment by performing 15 unperturbed stand-up motions. These trials revealed near-straight COM trajectories. Both the Trajectory Area and Initial Trajectory Area of these trials were consistently close to zero.

In our experiment, failures were an integral part of motor adaptation. When perturbation was first introduced in the perturbation block, none of the participants were

able to successfully perform the stand-up motion. The perturbation force pulled the participants backwards, which caused postural instability and required them to make at least one backward step to prevent falling over. Due to this backward motion, the movement trajectory of the early adaptation trials shifted in the direction of perturbation, resulting in a negative Trajectory Area. On average, the participants made 6.15 ± 3.10 failures before they could successfully perform the first stand-up motion.

To successfully perform the stand-up motions, and compensate for the perturbation force, the participants were required to adapt their usual stand-up motion by leaning against the anticipated perturbation. Even after their first successful trial, failures still sporadically occurred throughout the adaptation process. The participants made an average of 24.35 ± 11.28 failures. The average number of failures significantly dropped in the early trials with an exponential learning rate of 15.7 trials and reached the plateau at trial P40.74 ± 19.53. The adaptation process was reflected in the gradual increase of both Initial Trajectory Area and Trajectory Area measures in the direction against perturbation. The Trajectory Area increase demonstrates the overall effects of motor learning, while the increase in Initial Trajectory Area shows the effect of motor planning. On average, the participants needed to perform 84.35 ± 11.28 perturbed trials to reach the required 60 successful trials. At the end of the learning period participants successfully learned to adapt to the perturbation, indicated by the disappearance of failed trials and the prevalence of successful trials.

A single catch trial followed the 60 successful trials and none of the participants could successfully perform the stand-up motion in this condition. The Trajectory Area of the catch trial considerably increased in the direction against perturbation. After the perturbation block, the participants finished the experiment by performing 15 de-adaptation trials. Most participants failed to successfully perform the stand-up motion in the first de-adaptation trial, indicating that the participants were not able to instantly readapt to the unperturbed condition, despite the fact that they were consciously aware of it. Finally, both Trajectory Area and Initial Trajectory Area gradually returned to the same levels as in the baseline block.

4.2.1 Ecological Control: Success-Efficient/Failure-Safe Learning Strategy

To interpret the role of failed motion execution in the process of overall adaptation, we investigated how stand-up motion was influenced by preceding failed trials compared to successful ones. The Trajectory Area changes were significantly greater after failed trials than after successful trials ($t(19) = 2.92$, $p = .009$; Figure 4.2B). There was also a difference in changes in Initial Trajectory Area between trials following failure and trials following success, however this difference was not statistically significant ($t(19) = 1.62$, $p = .122$, Figure 4.2C). Specifically, after a failed trial, the COM trajectory bent in the direction against perturbation compared to the COM trajectory before a failed trial (Figure 4.2D). This resulted in an Initial Trajectory Area increase of the following successful trial, with respect to the successful trial performed prior to the failure ($t(19) = 2.23$, $p = .038$). Similarly, the Trajectory Area of the successful trial following a failed trial increased in

respect to the successful trial prior to the failure ($t(19) = 3.20$, $p = .005$). This corresponds to a change towards a safety-oriented COM trajectory where postural stability is maximised to prioritize safety. In this case, the trajectory is bent in the direction against perturbation, so the Trajectory Area and Initial Trajectory Area are both larger than those associated with an efficiency-oriented COM trajectory. On the other hand, the COM trajectory straightened during consecutive successful trials (Figure 4.2A). This resulted in Trajectory Area and Initial Trajectory Area decreases during the consecutive successful trials, though neither change was statistically significant ($t(19) = 1.22$, $p = .239$, $t(19) = 1.57$, $p = .134$). The direction of this effect corresponds to a shift towards an efficiency-oriented COM trajectory, where energy of motion is minimised. In this case trajectory is straightened and both the Trajectory Area and Initial Trajectory Area are reduced relative to the safety-oriented COM trajectory. Thus, overall, the participants adapted their movement patterns towards safety-oriented or energy-oriented, depending on whether they experienced a failure or a success.

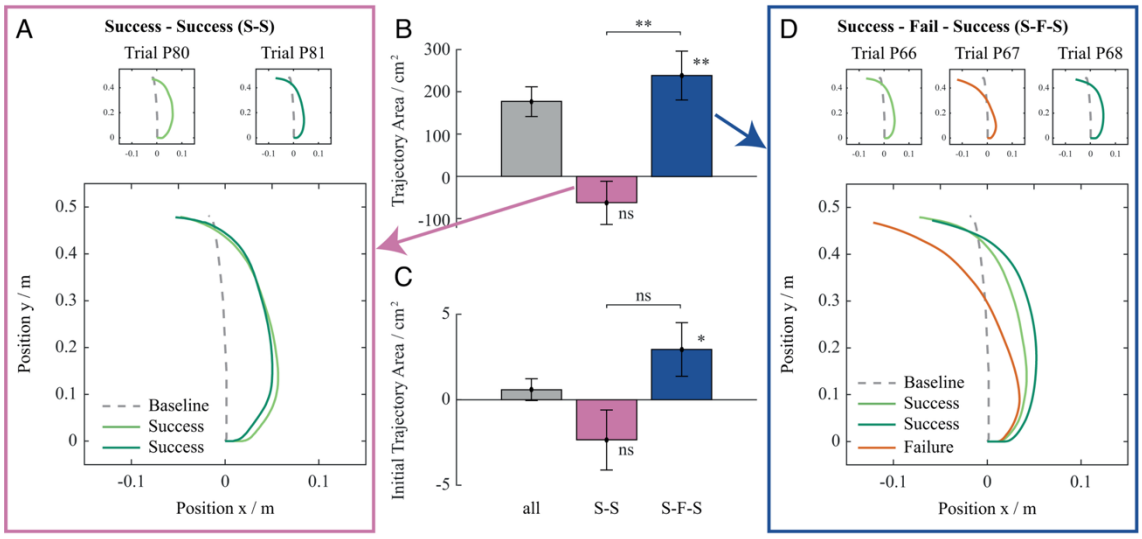


Figure 4.2: Representative centre of mass trajectories and adaptation metrics. (A) Representative set of centre of mass trajectories of two consecutive successful trials. Total changes of Trajectory Area (B) and Initial Trajectory Area(C), with changes between two consecutive successful trials (S-S) shown in pink and those between two consecutive successful trials separated by at least one failed trial (S-F-S) shown in blue. Error bars show standard error of the mean. (D) Representative set of centre of mass trajectories before and after a failed trial with successful trials shown in green and failed trials shown in red.

To further emphasize the strong effect of failed trials beyond their overall contribution to adaptation, we analysed the average changes in Initial Trajectory Area and Trajectory Area across different trial sequences: between two consecutive successful trials, between two consecutive successful trials separated by at least one failure, and between consecutive failed trials. On average, during consecutive successful trials, the Initial Trajectory Area and Trajectory Area decreased by -0.05 ± 0.04 cm² and -1.51 ± 1.14 cm², respectively. In contrast, in a successful trial that followed a failure, the Initial Trajectory Area and

Trajectory Area increased by 0.37 ± 0.18 cm² and 26.07 ± 8.59 cm², respectively, compared to the last successful trial before the failure. For consecutive failed trials, the average per-trial change was 0.53 ± 0.13 cm² for Initial Trajectory Area and 25.03 ± 10.00 cm² for Trajectory Area, reinforcing the role of failures in driving adaptation.

4.2.2 Motor Plan Learning

We observed quicker learning of the motor plan compared to the effects of the overall motor learning. This is reflected in a different rate of adaptation between the Initial Trajectory Area and Trajectory Area ($t(19) = 2.28$, $p = .034$). In particular, the Initial Trajectory Area significantly increased in early trials, with an exponential learning rate of 12.6 trials and a plateau at trial P34.35 ± 27.89 (black line in Figure 4.3C). The Trajectory Area increased over a longer period of time, with an exponential learning rate of 20.3 trials, and a plateau at trial P50.61 ± 23.90 (black line in Figure 4.3B). This shows that motor plan adaptation was completed while the overall adaptation to the perturbation was still ongoing. The motor plan no longer changed after the failures significantly decreased, while

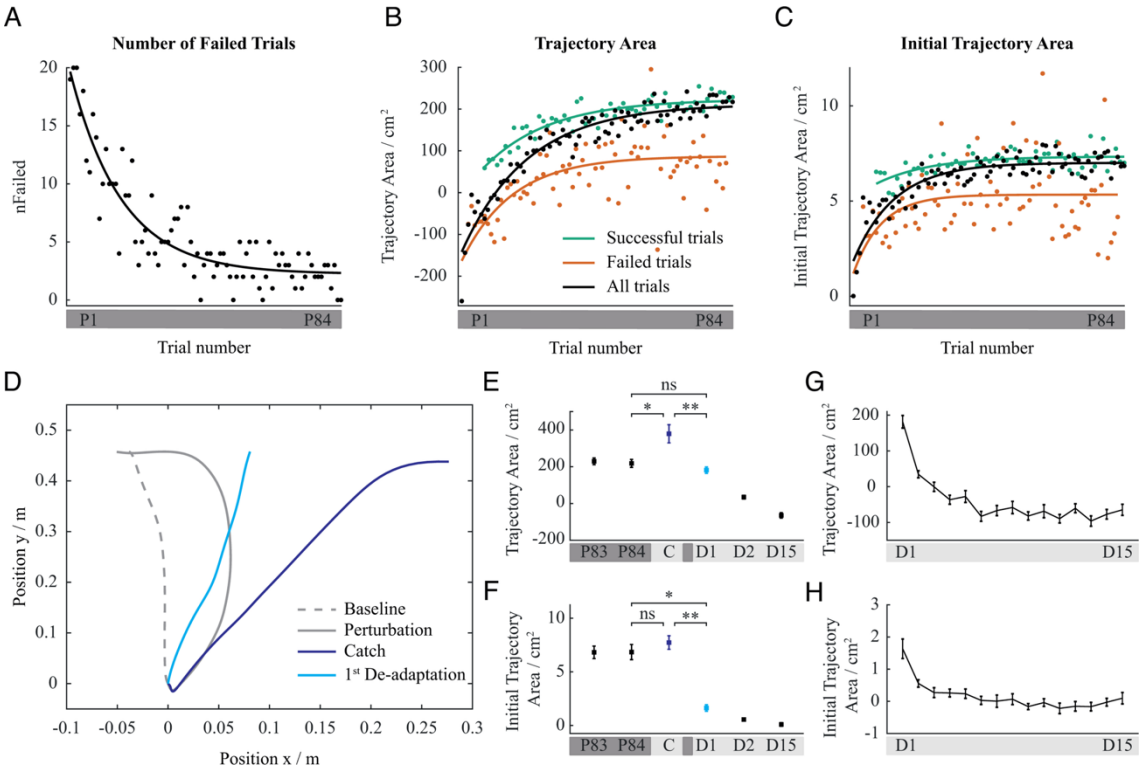


Figure 4.3: Adaptation dynamics during perturbation and de-adaptation. (A) The number of participants among the total of 20 that failed at each trial. Progression of Trajectory Area (B) and Initial Trajectory Area (C) during adaptation to perturbation where red represents failed trials, green represents successful trials, and black represents all trials. (D) Representative set of centre of mass trajectories for unperturbed trial, for last perturbed trial, for catch trial and for first de-adaptation trial. Comparison of Trajectory Area (E) and Initial Trajectory Area (F) of last perturbed trial, catch trial, and first de-adaptation trial. Progression of Trajectory Area (G) and Initial Trajectory Area (H) during the de-adaptation block. Error bars show standard error of the mean.

the adaptation of the whole COM trajectory continued even when failures became less frequent. Moreover, the motor plan only changed during failed trials and remained largely unchanged during successful trials. The Initial Trajectory Area increased during failed trials, with an exponential learning rate of 8.6 trials and a significant difference between the first and the last trials ($t(19)$ = -5.83, $p < .001$; red line in Figure 4.3C). It remained unchanged during successful trials, with no significant difference between the first and the last trials ($t(19)$ = -1.69, $p = .107$; green line in Figure 4.3C). On the contrary, overall adaptation was pronounced during both failed and successful trials. The Trajectory Area increased during failed and successful trials, as indicated by a significant difference between the first and the last trials (failed trials: $t(19) = 7.14$, $p < .001$, red line in Figure 4.3B; successful trials: $t(19) = 5.37$, $p < .001$, green line in Figure 4.3B). The exponential learning rates were 15.6 and 20.3 trials, respectively. Moreover, quicker effects of motor plan learning compared to the effects of the overall motor learning were also observed in the de-adaptation period. In particular, more trials were necessary for the Trajectory Area to return to baseline values than for the Initial Trajectory Area ($t(19) = 3.97$, $p < .001$; Figure 4.3G and H). Specifically, the Trajectory Area reached the baseline plateau after 12.48 ± 3.79 trials, while the Initial Trajectory Area required only 7.13 ± 4.86 trials.

To further elucidate the mechanism of high-level motor planning, we compared the motion during the catch trial with the motion during the first de-adaptation trial (Figure 4.3D). Specifically, this allowed for an assessment of the effect of participant awareness of upcoming perturbation-free trials. Participants made no change to their motor plan when they were unaware that the perturbation will be absent. However, when they were informed about the absence of the perturbation, their motor plan changed. There was a significant difference in the Initial Trajectory Area between the last perturbed trial, the catch trial and the first de-adaptation trial ($F(2, 38) = 10.57$, $p < .001$). Specific pairwise comparisons revealed that the Initial Trajectory Area of the catch trial was significantly larger than that of the first de-adaptation trial ($t(19) = 8.78$, $p < .001$). The Initial Trajectory Area of the catch trial was similar to the preceding perturbed trial ($t(19) = 0.37$, $p = .716$), while it significantly decreased during the first de-adaptation trial ($t(19) = 10.14$, $p < .001$), almost reaching the level of Initial Trajectory Area during unperturbed motion (Figure 4.3F). Consequently, there was also a difference on the level of the whole trajectory between the perturbed, catch and first de-adaptation trials ($F(2, 38) = 3.41$, $p = .044$). This is reflected in a significant difference of the Trajectory Area between the catch and first de-adaptation trial ($t(19) = 3.69$, $p = .0015$). Trajectory Area of the catch trial increased with respect to the preceding perturbed trial ($t(19) = 2.77$, $p = .012$), while it remained approximately the same during the first de-adaptation trial ($t(19) = 1.89$, $p = .074$; Figure 4.3E). Moreover, the relative reduction in the first de-adaptation trial with respect to the catch trial was significantly larger for the Initial Trajectory Area (0.69 ± 0.34) than for Trajectory Area (0.33 ± 0.29) ($t(19) = 4.59$, $p < .001$). This demonstrates quicker changes on the level of the motor plan compared to the changes on the level of the whole trajectory.

4.2.3 Internal Model Learning

To understand the adaptation process distinctive from the adaptation of the plan, we further investigated the adaptation of the lower-level feed-forward motor control. Internal model changes during the adaptation to perturbation are reflected in aftereffects during catch and de-adaptation trials. None of the participants could successfully perform the stand-up motion during the catch trial, and most participants failed to successfully perform the stand-up motion in the first de-adaptation trial. The Trajectory Area of the catch trial was considerably larger than the Trajectory Area of the perturbed trials (Figure 4.3E). Moreover, the Trajectory Area of the first de-adaptation trial was larger compared to the unperturbed trials, despite the Initial Trajectory Area of this trial being noticeably smaller than the Initial Trajectory Area of the preceding perturbed trials, and similar to the Initial Trajectory Area of unperturbed trials (Figure 4.3F). Nevertheless, despite the motor plan during the de-adaptation trial being changed appropriately with respect to no anticipated perturbation, participants still mostly failed to successfully perform the stand-up motion. Improvement of the internal model during the adaptation to perturbation is further reflected in a progressive increase of Smoothness over the whole perturbation block with an exponential learning rate of 29.4 trials (black line in Figure 4.4A), and a plateau at trial P58.13 ± 26.11. This is significantly slower than the reduction of the number of failed trials

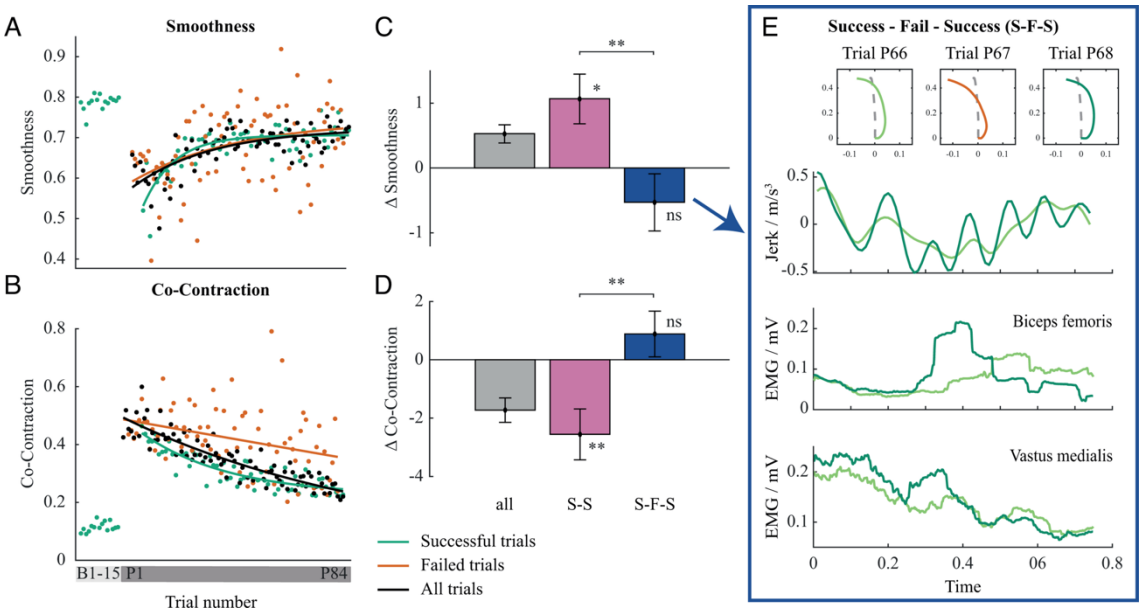


Figure 4.4: Progression of Smoothness and Co-Contraction. The progression of Smoothness (A) and Co-contraction (B) during adaptation to perturbation, with red representing failed trials, green representing successful trials, and black representing all trials. Total changes of Smoothness (C) and Co-Contraction (D) where change between two consecutive successful trials (S-S) is shown in pink and change between two consecutive successful trials separated by at least one failed trial (S-F-S) is shown in blue. (E) Representative jerk of centre of mass trajectory and EMG trajectories for successful trials before a failed trial (light green) and after a failed trial (dark green). Error bars show standard error of the mean.

($t(15) = 4.61$, $p < .001$) and the related adaptation of the motor plan ($t(15) = 2.96$, $p = .010$). Moreover, Smoothness and its rate of change is comparable for both successful and failed trials (Figure 4.4A), indicated by no statistically significant difference in Smoothness between the two ($t(19) = 1.01$, $p = .324$).

4.2.4 Feedback Gains are Manipulated Based on Success and Failure

In line with internal model improvements, a concomitant decrease in feedback contribution was observed. Co-Contraction plateaued at trial P67.42 ± 17.70, which was not significantly different to the plateau of Smoothness ($t(11) = 1.43$, $p = .179$). The statistical comparisons of when adaptation measures reached plateaus are shown in Table 4.1. Although the contribution of feedback steadily decreased during the adaptation to perturbation, this was notable only during consecutive successful trials. Feedback contribution during failed trials remained high. Co-Contraction of failed trials throughout the adaptation to perturbation was high (red line in lower panel of Figure 4.4B), there was a significant difference between the mean Co-Contraction of successful and failed trials ($t(19) = 3.22$, $p = .005$), and no significant difference of Co-Contraction between the first failed and last failed trials ($t(19) = 1.32$, $p = .204$).

Table 4.1: Statistical comparison of adaptation plateaus.

	Co-Contraction	Smoothness	Trajectory Area	Initial Trajectory Area
Number of Failed Trials	**$t = 3.40$ $p = .004$ ****	**$t = 4.61$ $p < .001$ *****	**$t = 2.19$ $p = .042$ ***	$t = 1.00$ $p = .328$
Initial Trajectory Area	**$t = 3.62$ $p = .002$ ****	**$t = 2.96$ $p = .010$ ***	**$t = 2.28$ $p = .034$ ***	-
Trajectory Area	$t = 1.99$ $p = .065$	$t = 1.82$ $p = .089$	-	-
Smoothness	$t = 1.44$ $p = .179$	-	-	-
Co-Contraction	-	-	-	-

Comparisons with significant differences are shown in bold. They are corrected for multiple comparisons using Benjamini-Hochberg adjustment. *** < .001, ** < .01, * < .05

Contribution of feedback decreased only during consecutive successful trials and it increased when a failure occurred. This is reflected in the statistically significant decrease in Co-Contraction during consecutive successful trials ($t(19) = 3.18$, $p = .005$). On the other hand, the Co-Contraction of a successful trial after failure increased compared to a

successful trial before a failure (Figure 4.4D). However, the change was not statistically significant ($t(19) = 1.23$, $p = .235$). Increase of feedback gains after failed trials, reflected by Co-Contraction increases, also caused the Smoothness to decrease. However, the change was not statistically significant ($t(19) = 1.20$, $p = .245$). On the other hand, there was a statistically significant increase of the Smoothness during the consecutive successful trials ($t(19) = 2.78$, $p = .012$; Figure 4.4C).

4.3 Discussion

We investigated whole-body motor adaptations where a risk of failure was an integral part of the learning process. In contrast to arm-reaching studies where failed trials are classified as trials when the participant did not successfully reach the desired target (Ikegami et al., 2021), failures in our experiment are movements when the participant lost postural balance and was required to make corrective steps to prevent a fall. In our study, a considerable number of failed trials were observed during learning. At the end of the learning period, participants successfully learned to adapt to the perturbation resulting in the elimination of failed trials and the prevalence of successful trials. This is in line with previous research where the number of failed trials decreased during arm-reaching (Ikegami et al., 2021) and whole-body motor learning (Babič et al., 2016).

In this study, safe movement refers to a centre-of-mass trajectory shifted in the direction opposite to the backward perturbation, thereby counteracting the perturbation and reducing the risk that the participant would exceed the limits of postural stability and require a corrective step. Efficient movement refers to a straighter centre-of-mass trajectory with smaller Trajectory Area and Initial Trajectory Area, representing less unnecessary deviation and a movement pattern intended to minimise energetic cost.

4.3.1 Balancing Safety and Efficiency

Motor behaviour was predominantly adapted after failed trials which clearly shows reinforcement learning mechanisms (Kawato & Samejima, 2007; Merel et al., 2019; Morimoto & Doya, 2001; Sutton & Barto, 1998). After the failed trials, the COM trajectory shifted in the direction against perturbation resulting in a shift towards a safer and more stable movement pattern. Conversely, consecutive successful trials led to a straightened trajectory and resulted in a more efficient movement pattern. The behavioural results together with the computational study strongly support that trial-based task outcome is responsible for higher level plan adaptation to maximise ecological fitness through reinforcement learning (Ito & Doya, 2011; Maia & Frank, 2011; Samejima et al., 2005; Seymour, 2019). Under this theory, failures push the plan towards safer execution of motion, while successful motions promote optimization of movement efficiency.

When changes in the environment are perceived as a potential danger during the execution of movements, risk-averse learning with prioritised safety takes place (Greenberg et al., 2022). This allows for the learning process to avoid the risk of injury and in the long term leads to higher chances of survival (Krakauer et al., 2017). Once the environment is no longer perceived as dangerous, cost functions that relate to the execution of the motion, such as minimization of effort, can be prioritized (Scott, 2004). Only cautious execution of

motion in dangerous situations (failure-safe) together with the minimization of effort in safe situations (success-efficient) can lead to truly optimised ecological fitness.

This success-efficient/failure-safe learning mode switching mechanism is similar to the win-stay/lose-shift strategy observed in neuroeconomics and social linguistics (M. A. Nowak & Sigmund, 2004; M. Nowak & Sigmund, 1993). The win-stay/lose-shift strategy is a heuristic learning strategy to retain a successful action but to switch to another action after a failure (Posch, 1999). In our experiment participants retained and optimised their movements during successful motions, but switched to a safer movement after a failure. By switching between learning modes based on success and failure, individuals were able to quickly adapt to new motor tasks while minimising the risk of injury.

From the perspective of the exploitation-exploration trade-off in human decision-making (Cohen et al., 2007) and reinforcement learning (Sutton & Barto, 1998), our model suggests that successful motions encourage the exploration of more energy-efficient movements, whereas failures lead to a switch to a safe policy that ensures task completion without failure, thereby functioning as exploitation. In humans, the exploration-exploitation trade-off is dynamically adjusted and task-specific (Averbeck, 2015; Shen & Dayan, 2024). This trade-off can be modulated by context (Komar et al., 2019) and emotional state (van Dooren et al., 2021).

In tasks where there is no risk of injury, exploration is typically favoured at the outset, while exploitation becomes the dominant strategy over time (Humphries, 2012; Shen & Dayan, 2024). However, in response to unexpected failure or loss, exploration levels tend to increase (Pekny et al., 2015; Shen & Dayan, 2024; van Mastrigt et al., 2020), which contrasts with our task. We argue that this is due to the risk of injury and the high energy demands of our task, distinguishing it from other contexts where failure does not pose a risk of harm.

Several studies have shown that tasks learnable through error feedback can also be acquired via reinforcement feedback alone (Cashaback et al., 2017; Izawa & Shadmehr, 2011; Therrien et al., 2016; Truong et al., 2023). From a computational standpoint, error feedback provides richer information than reinforcement feedback and should be leveraged whenever available. This observation is supported by experimental findings (e.g., (Izawa & Shadmehr, 2011), although mixed results exist (Cashaback et al., 2017). These studies also confirm that the central nervous system employs multiple learning strategies to solve sensorimotor learning tasks, and that they are recruited in a task-dependent way. It has been also showed that combining reinforcement and error-based learning improves skill acquisition and retention in the full-body movement of basketball shooting (Truong et al., 2023). However, the interplay between these learning mechanisms and the task-dependent recruitment of reinforcement learning processes in the brain remains poorly understood (D. Lee et al., 2012; Therrien et al., 2016). We argue that these two learning processes operate hierarchically under ecological constraints, yet run independently and in parallel, enabling rapid responses essential for ecological fitness and survival.

In the complex task of controlling the body for safe movement under altered dynamics, the central nervous system must simultaneously learn to represent new body-environment dynamics while avoiding failure. The former is supported by prediction error feedback, which is continuously processed through sensorimotor loops in the central nervous system.

The latter, however, relies on delayed feedback, only available at the end of a movement in the form of success or failure, necessitating reinforcement learning. Experimental evidence suggests that feedback error learning is mediated by the cerebellum, while reinforcement learning involves the basal ganglia and multiple brain structures encoding value functions (Doya, 2000; D. Lee et al., 2012), indicating a division of labour and modularity in the brain.

4.3.2 Modularity and Hierarchy in Motor Learning

We observed faster adaptation in motor plan learning compared with overall motor learning. While motor plan learning took place in early trials and was no longer changed after the failures significantly decreased, overall motor learning continued over a longer period of time even after the failures were no longer frequent (as in (Ikegami et al., 2021)). Similarly, in the de-adaptation period, the motor plan was more quickly changed, while it took longer for the overall trajectory to return to baseline values. This shows two distinctive plan and motor control processes with different rates of adaptation, faster on the level of planning and slower on the level of internal model learning, which indicates motor control mechanism modularity. This is in line with observations that motor adaptation processes in the brain rely on distinct neural systems that operate on different timescales and retain information at different rates (Fusi et al., 2007; J.-Y. Lee & Schweighofer, 2009; Smith et al., 2006; Yang & Lisberger, 2010). In particular, slow mechanisms retain motor memories on longer times scales, whereas fast mechanisms enable us to quickly adapt to potential transient changes in the environment (Fusi et al., 2007; Wolpert & Flanagan, 2010). Moreover, in our study, the motor plan significantly changed only during failed trials. This suggests that only failed execution of motion activates the adaptation mechanism to adapt the plan in order to maximise ecological fitness. On the contrary, the whole trajectory changed both during failed and successful trials. This further demonstrates that internal model adaptation is distinctive from the adaptation of the plan and is governed by optimality criteria not directly related to ecological fitness (e.g., minimization of motor prediction error). Previous work suggests that the fast and slow processes of learning may be underpinned by explicit strategy learning and implicit motor adaptation (McDougle et al., 2015; Taylor & Ivry, 2014). Moreover, Ikegami et al. (Ikegami et al., 2021) have proposed that failure driven adaptation is an explicit strategy learning since it is faster than implicit components represented by internal model adaptation and is only active in the presence of failures (McDougle et al., 2015; Schween et al., 2020). Explicit strategy learning enables us to quickly adjust the motor plan to a suitable solution which ensures the safe execution of motion, while the optimization process persists (McDougle et al., 2016).

To further elucidate the mechanisms and the relation between the two distinctive adaptation processes, we compared the motion during the catch trial with the first de-adaptation trial. This allowed us to assess the influence of anticipation, as participants were first unaware that the perturbation would be turned off (catch), and later informed that there would then be no perturbation (de-adaptation). Based on the previous arm-reaching studies (Kluzik et al., 2008), we would expect that despite the verbal instructions, participants would exhibit large aftereffects during the de-adaptation trial comparable in

magnitude with the aftereffects of the catch trials. However, as our results indicate, the motor plan was significantly different between the two trials. While the motor plan for the catch trial remained similar to the preceding perturbed trials, it significantly changed during the first de-adaptation trial when the participants consciously made quick explicit changes to the motor plan as they were aware that the perturbation would be absent. A possible reason for the discrepancy between our results, and the results of previous arm-reaching studies, could lie in the ecological aspects of our experiment. Specifically, the cost of failure during the arm-reaching movement is negligible, but very pronounced in our whole-body movement where a failure would carry a substantial injury risk. Nevertheless, although the motor plan during the de-adaptation trial was changed appropriately with respect to absence of perturbation, participants still mostly failed to successfully perform the stand-up motion. This suggests that participants were not able to instantaneously re-adapt the internal model as they did for the motion plan. In this regard, our results are in line with previous work suggesting that implicit learning occurs without conscious awareness and can result in large and prolonged aftereffects which reflect changes to the internal model during previous adaptation (Miyamoto et al., 2020). This implies hierarchical organisation of the two adaptation processes. The motor plan was adapted on a higher level by conscious understanding of the motor task while on the lower level, the internal model could not be consciously changed and needed more time to re-adapt.

4.3.3 Interplay of Internal Model Adaptation and Feedback Contributions

Prominent aftereffects during catch and de-adaptation trials reflect adaptation on the level of internal model of movement dynamics. The aftereffects of force adaptation have been interpreted as evidence that the nervous system learns to anticipate and counteract novel forces by building an internal model of the body and the environment (Kawato, 1999; Kluzik et al., 2008; Miall & Wolpert, 1996; Shadmehr, 2004; Thoroughman & Shadmehr, 2000; Wolpert et al., 1995). Improvement of the internal model during the adaptation to perturbation is further evident in a progressive increase of Smoothness over the whole perturbation block (Huang & Shadmehr, 2009). This suggests that, alongside the refinement of the motor plan, the internal model becomes more accurate in predicting and guiding smoother movements. The observed increase in Smoothness was slower than the rate of error reduction, indicating a more incremental refinement of the internal model (Izawa et al., 2008). Notably, the improvement of the internal model, which relies on time-sensitive sensory feedback, exhibits similar patterns in both successful and failed trials. This suggests that the internal model undergoes adaptive changes during both successful and failed attempts at motion execution. These results suggest that the adaptation of the internal model may occur at lower levels of motor control and could be independent of reinforcement learning mechanisms.

In contrast with the improvement of the internal model during both successful and failed motion, feedback contribution of the motor control decreased only during the successful motion execution. On the other hand, motor control feedback contribution remained high and largely unchanged during the failed trials, despite the improvement of

the internal model. The effect of a failed trial was notable even in the subsequent successful motion, where the feedback contribution increased with respect to the successful motion before the failed trial. This could happen since, due to the recent failure, the internal model was not trusted so feedback contribution was amplified to increase robustness to perturbation and to successfully perform the motion (S. Franklin et al., 2012; Osu et al., 2002; Tee et al., 2010).

4.3.4 Limitations

The relationship between co-contraction, impedance control, and feedback control is critical for understanding motor adaptation mechanisms. Co-contraction is primarily associated with impedance control, which modulates joint stiffness and stability to handle perturbations (e.g., Burdet et al., 2001; Franklin et al., 2008). Feedback control, on the other hand, involves real-time adjustments to motor output based on sensory input errors, often modeled as part of an optimal control framework (Todorov & Jordan, 2002). While co-contraction can indirectly influence feedback mechanisms through automatic gain scaling (Pruszynski et al., 2009), it is not a direct proxy for feedback control. In this study, co-contraction was used as an indirect measure to infer changes in feedback gains rather than feedback control itself, leveraging prior evidence that feedback gains are scaled by muscle activity (Pruszynski et al., 2009) and that co-contraction enhances feedback responses (Akazawa et al., 1983; Doemges & Rack, 1992; Krutky et al., 2010; Maurus et al., 2024). However, this interpretation omits the specific role of impedance control mechanisms that arise from co-contraction. As demonstrated by Calalo et al. (Calalo et al., 2023), the sensorimotor system modulates co-contraction relative to visuomotor feedback responses to regulate movement variability. Their findings underscore that co-contraction primarily functions to increase joint stiffness and reduce movement variability, whereas feedback control dynamically corrects for positional deviations. However, unlike our study, they measured visuomotor feedback responses directly, which allowed them to explore the relationship between co-contraction and feedback responses with greater precision. Our study lacks this direct assessment, limiting our ability to comprehensively characterize the interplay between these two mechanisms.

4.4 Conclusion

Our findings show that human whole-body adaptation is hierarchically organised. At the top level, the ecological control layer determines what should be optimised during motor task learning by specifying the reward/cost function used by reinforcement learning to form motor plans to be used by lower-level control. In turn, the motor-control processes include internal model learning and feedback gain tuning mechanisms in addition to the feedback control to achieve the desired plan. As such, both motor-planning and motor-control processes contribute to adaptation but at different rates and based on different optimality criteria. The motor plan is adapted quickly through reinforcement learning mechanisms after failed motion execution. Conversely, motor control adaptation is slower, and predominantly occurs as a result of internal model adaptation, irrespective of whether the movement proves to be successful.

Based on the results of this study, we argue that the optimality concept in computational motor control should be redefined from an evolutionary viewpoint where ecological fitness plays the key role in generating movements, integrating injury risk in addition to cost functions that prioritise movement.

4.5 Methods

4.5.1 Participants

Twenty healthy male participants (age 23.1 ± 2.1 (SD) years; height 1.83 ± 0.06 (SD) m; mass 77.00 ± 9.97 (SD) kg) participated in the study. An informed written consent form, approved by the National Medical Ethics Committee of Slovenia (No. 0120-339/2017/7), was signed by all participants prior to their participation. The experimental protocol was approved by the National Medical Ethics Committee of Slovenia (No. 0120-339/2017/7) and carried out in accordance with the Oviedo Convention.

4.5.2 Experimental Procedure

Participants were asked to perform a series of squat-to-stand motions (trials) as depicted in Figure 4.1A. Movements of the foot, shank, thigh, and torso were recorded with the motion capture system Optotrak 3D Investigator (NDI, Waterloo, Ontario) in real time with a sampling rate of 1000 Hz. We firmly attached ten active markers on the apparent axis of rotation of the left and right fifth metatarsal joints, ankles, knees, hips and shoulders. Two additional markers were placed on the back, on the spinous processes of the 1st lumbar - L1, and 1st thoracic - T1 vertebra, to measure trunk angle.

The trial started from a squat position with the trunk angle between 0° and 10° in the forward direction with respect to the vertical axis. To ensure consistency of the starting squat positions, a screen was placed in front of the participants displaying the trunk angle and the height of the L1 marker. The trial ended when participants reached a fully extended standing position. The participants were asked to perform the squat-to-stand motions in rhythm with the metronome set to 100 bpm. To avoid fatigue, there was a 30 s pause in between each trial. Following visual inspection, we excluded 1.7% of the data due to atypical motions exhibited by the participants. A trial was considered a success if the participants reached the standing position without lifting their feet, and a failure if the participants needed to make a corrective step to keep their postural balance. All successful and failed trials were recorded and analysed. To accommodate inter-individual variability in learning rates of individual participants, we normalised the number of perturbed trials prior to the catch trial to 84. Specifically, for a participant with N perturbed trials, we computed the normalized trial index as (trial number / N) × 84.

Experiments were performed using a custom-made force-controlled pulling mechanism which generated perturbations in the backward direction. Perturbations were applied at the waist, which allowed for a direct perturbation of the participant's centre of mass (COM). The vertical velocity of the participant's COM was used to calculate the force of perturbation F_{pert} as

$$F_{pert} = \mathrm{K} \cdot \dot{y}_{COM} \cdot m_{sub}, \tag{4.1}$$

where $\dot{y}_{COM}$ is the velocity of the participant's COM in the upward direction, m_{sub} is the mass of the participant, and the parameter K is a constant with a value of 3 s^{-1}. The constant K was experimentally determined in a preliminary squat-to-stand experiment and was set to a value that generated a perturbation at which untrained participants needed to make a corrective step to keep their postural balance. COM was calculated by using a reduced three-dimensional biomechanical model of the human body. The model consists of five body segments (left/right shank and thigh and the trunk) with their inertial properties based on specifications from De Leva (de Leva, 1996).

4.5.3 Measures and Statistical Analysis

To quantify motor processes, we used four measures: Trajectory Area, Initial Trajectory Area, Smoothness and Co-Contraction. Due to the different anthropometric parameters of individual participants, and due to the fact that individual trials had slightly different starting and ending points in spite of the visual feedback and the COM target specification, the trajectories were translated to start from the common origin and y was scaled to 0.5 m. This translation was for computational convenience and did not affect the statistical results.

The Trajectory Area represented the total deviation of the COM trajectory with respect to the straight line in the sagittal plane originating from the start COM position. The Initial Trajectory Area represented the COM trajectory deviation with respect to the straight line of the first 5% of the movement, which on average occurred within 96.4 ± 7.0 (SD) ms from the start of the movement and approximately corresponded to 2.5 cm in height. Area, either Trajectory Area or Initial Trajectory Area, of a given perturbed trajectory *T(x(t), y(t))*, was defined as the path integral of the distance of the trajectory points to the straight line in the sagittal plane:

$$A = \int_{y_{start}}^{y_{end}} x(t)dy, \tag{4.2}$$

where y_{start} is a starting COM height in the squat position, and y_{end} is 2.5 cm for the Initial Trajectory Area and COM height in the standing position for the Trajectory Area. Positive values for Trajectory Area and Initial Trajectory Area represent the forward direction (direction against the perturbation) while negative values represent backward direction (the direction of perturbation).

Smoothness was calculated as a dimensionless squared jerk (Hogan & Sternad, 2009)

$$SM = \left(\int_{t_1}^{t_2} \dddot{x}(t)^2 dt \right) D^3 / v_{mean}^2, \tag{4.3}$$

where D is the duration and v_{mean} is the mean speed of the movement.

Co-Contraction of the thigh antagonistic pair of muscles was calculated as a quotient of integrated muscle activity of vastus medialis $iEMG_{vm}$ and biceps femoris $iEMG_{bf}$ (Falconer & Winter, 1985)

$$SM = \frac{\left(\int_{t_1}^{t_2} \dddot{x}(t)^2 dt\right) D^3}{v_{mean}^2}. \tag{4.4}$$

The EMG signals were recorded at a sampling rate of 1000 Hz using surface electrodes (SX230-1000, Biometrics Ltd., Newport, UK). The recordings were performed synchronously with the motion capture system. Both EMG signals were band-pass filtered using a zero-lag, 4th-order Butterworth filter with cut-off frequencies of 20 and 500 Hz. The signals were then full-wave rectified, and root mean square (RMS) envelopes were computed. Finally, the RMS envelopes were integrated over time (iEMG) to express the magnitude of muscle activity.

For all four measures we used paired t-tests to compare the total change of the values between two consecutive successful trials, and between two consecutive successful trials separated by at least one failed trial in between. For each of the two conditions, we first calculated the sum of all changes during the perturbation block for each participant and then calculated the average values across all participants. This approach was chosen to capture how failed trials uniquely influence motor adaptation by comparing changes in successful movements with and without intervening failures. Since failed motions are significantly different from successful motions, directly comparing the two would primarily reflect their inherent biomechanical differences rather than the influence of failures on motor adaptation. By focusing on changes between successful trials with and without failures in between, we isolated the specific contribution of failed trials to the adaptation process. Failed trials consisting of a forward step – opposite to perturbation – represented 1.4% of all failures. These trials were excluded from subsequent analysis. Moreover, we also calculated an average of the total change of the measures between the first and the last trial in the perturbation block. We also used paired t-tests to test if Initial Trajectory Area and Trajectory Area changed significantly from one successful trial to the next. This was tested separately for the case of consecutive successful trials, and consecutive successful trials with at least one failed trial in between.

We determined the learning rates for the average values of Initial Trajectory Area, Trajectory Area, Smoothness, Co-Contraction, and the number of failed trials using an exponential curve $a \cdot e^{-\frac{x}{b}} + c$ that was fitted to average data, where x is the trial index, and the time constant b represents the learning rate. Moreover, for each participant, we identified the plateau trial at which these metrics stabilised, defined as reaching 95% of their value on the fitted exponential curve. Statistical analyses were then conducted based on these plateau values. In Co-Contraction fitting, 3 participants were excluded due to their noisy EMG data, and 1 participant was excluded due to the negative value of the time constant b. In Smoothness fitting, 3 participants were excluded due to a linear fit with infinite learning rate, and 1 participant was excluded due to the negative value of the time constant b. To determine if the Trajectory Area or Initial Trajectory Area of successful

trials increased during the adaptation period, we compared the average values of the first and last three trials with a paired t-test.

To compare Initial Trajectory Area and Trajectory Area between catch trial, the 1st de-adaptation trial, and last perturbed trial, we performed repeated measures ANOVA and additional post-hoc paired t-tests with Bonferroni correction. The relative reductions in the Trajectory Area and Initial Trajectory Area in the first de-adaptation trial with respect to the total change due to the preceding adaptation, was calculated as

$$RD = 1 - \frac{A_{D1} - A_B}{A_C - A_B}, \tag{4.5}$$

where A is either Initial Trajectory Area or Trajectory Area, and the subscripts $D1$, C, and B denote the first de-adaptation trial, catch trial, and baseline trial, respectively. Relative reductions in Trajectory Area and Initial Trajectory Area were compared with a paired t-test.

Chapter 5

Computational Model Integrating Sensorimotor and Reward Learning

5.1 Introduction

To further explain how sensorimotor adaptation and explicit reward learning are combined to drive motor adaptations we designed a computational model that combines both motor learning processes. Even though we can separate the sensorimotor adaptation and explicit reward learning processes in human motor experiments fairly well, we cannot guarantee a perfect separation due to the inability to fully isolate the two processes or completely inhibit one of them. For this reason, we investigated the individual contributions of either sensorimotor adaptation or explicit reward learning on its own by simulating them separately with a computational model. We built an iterative learning control model with modules that differentiate between sensory and reward prediction errors. This allowed us to get an insight into how these two learning processes are integrated into motor control. The model successfully replicates the experimental results, validating that it effectively captures these learning processes. Moreover, the sensorimotor adaptation and explicit reward learning processes explain the differences in aftereffects and the ability to switch between perturbations.

5.2 Iterative Learning Control Computational Model

To investigate the individual characteristics and contributions of either sensorimotor adaptation or explicit reward learning to the overall motor learning, we designed a computational model which distinguishes between the two learning processes. By using the simulation, we were able to analyse sensorimotor adaptation and explicit reward learning processes separately and make comparisons between their outcomes. We modelled the experimental system of a human arm reaching towards a target where the arm was represented as a point mass. The system state represented position and velocity in either a polar coordinate system or a cartesian coordinate system for visuomotor and force-field adaptation respectively. Performance of the system was determined by the difference between the desired trajectory and actual trajectory and included errors in both position and velocity. To model the human learning process, we used an iterative learning control

(ILC) system, where a current-iteration feedback controller is incorporated with ILC in the parallel architecture and combined with learning based on reward prediction error. We first designed an ILC model in which changes are made directly to the motor commands. Afterwards, we designed an adaptive control model based on the same concept as the simple ILC model, but with changes applied at the controller level.

5.2.1 ILC Model for Visuomotor Rotation

To simulate adaptation in the visuomotor rotation paradigm, we first implemented ILC model that makes changes directly to the motor commands as shown on Figure 5.1. The goal of this model was to reproduce the behavioural dynamics observed in the reaching task under rotated visual feedback.

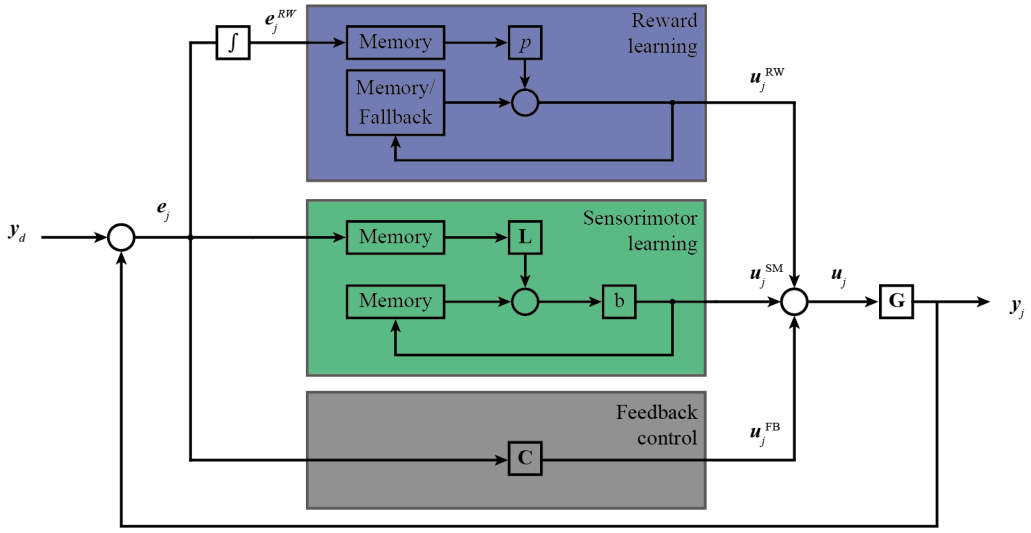


Figure 5.1: Architecture of the computational model. The parallel architecture computes the control signal $\boldsymbol{u}_j$ as a sum of control signals that are based on explicit reward learning $\boldsymbol{u}_j^{\mathrm{RW}}$, sensorimotor adaptation $\boldsymbol{u}_j^{\mathrm{SM}}$ and feedback controller $\boldsymbol{u}_j^{\mathrm{FB}}$. It directly generates the motor commands to the plant G. The integrator block ∫ represents the calculation of the reward error $\boldsymbol{e}_j^{\mathrm{RW}}$ which can be end-point error or binary information. Fallback module represents the initial conditions of motor commands for the first simulation trial and memory module represents motor commands from the previous simulation trial.

We modelled the experimental system of a human arm reaching towards a target as seen in the experimental study. The human arm was modelled as a point mass with the mass $\boldsymbol{m}$ and described in polar coordinate system with the radial distance $\boldsymbol{d}$ and polar angle γ with respect to the initial position. Our system was designed as a discrete-time, linear time-invariant multiple inputs multiple outputs (MIMO) system

$$\boldsymbol{x}_j(k+1) = \mathbf{A}\boldsymbol{x}_j(k) + \mathbf{B}\boldsymbol{u}_j(k) \tag{5.1}$$

$$\boldsymbol{y}_j(k) = \boldsymbol{x}_j(k) + \boldsymbol{r}_j, \tag{5.2}$$

where $k \in \{1, \ldots, k_{end}\}$ is the time step and $j \in \{1, \ldots, j_{end}\}$ is the simulation trial index. Moreover, $\boldsymbol{u}_j(k)$ is the control vector

$$\boldsymbol{u}_j(k) = \begin{bmatrix} f_j^d(k) \\ f_j^\gamma(k) \end{bmatrix},$$

with $f_j^d(k)$ being the control force in the radial direction and $f_j^\gamma(k)$ in the polar angle direction. $\boldsymbol{x}_j(k)$ is the state vector defined as

$$\boldsymbol{x}_j(k) = \begin{bmatrix} d_j(k) \\ \dot{d}_j(k) \\ \gamma_j(\mathrm{k}) \\ \dot{\gamma}_j(k) \end{bmatrix},$$

and the output $\boldsymbol{y}_j(k)$ is defined similarly. $\boldsymbol{r}_j$ is a rotation vector

$$\boldsymbol{r}_j = \begin{bmatrix} 0 \\ 0 \\ \beta \\ 0 \end{bmatrix},$$

where β is the angle of rotation that can either be 0 when there is no visuomotor rotation or ±30° for clockwise and counter clockwise visuomotor rotation. The state matrix is defined as

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and the input matrix as

$$\mathbf{B} = \begin{bmatrix} 0 & 0 \\ 1/m & 0 \\ 0 & 0 \\ 0 & 1/md^2 \end{bmatrix}.$$

Performance of the system is determined by the error vector $\boldsymbol{e}_j(k)$ in both position and velocity:

$$\boldsymbol{e}_j(k) = \begin{bmatrix} e_j^d(k) \\ e_j^{\dot{d}}(k) \\ e_j^\gamma(\mathrm{k}) \\ e_j^{\dot{\gamma}}(k) \end{bmatrix}.$$

It is defined as a difference between the desired trajectory and the actual trajectory and is calculated as

$$\boldsymbol{e}_j(k) = \boldsymbol{y}^d(k) - \boldsymbol{y}_j(k) + \boldsymbol{n}_\mathrm{s}, \tag{5.3}$$

where $\boldsymbol{y}^d(k)$ is the desired trajectory, $\boldsymbol{y}_j(k)$ is the actual trajectory and $\boldsymbol{n}_s$ is a normally distributed sensory noise with variance $\sigma_d^2 = 0.002$ for distance, $\sigma_\gamma^2 = 0.3$ for angle, $\sigma_{\dot{d}}^2 = 0.004$ for velocity and $\sigma_{\dot{\gamma}}^2 = 0.05$ for angular velocity. End-point error for explicit reward learning is calculated as

$$\boldsymbol{e}_j^{\mathrm{RW}} = \sum_{l=1}^{k_{end}} (2e_j^{\gamma}(l) \cdot d_j(l)) / k_{\mathrm{end}} \tag{5.4}$$

and is the same for all time steps. In cases where only end-point information is provided and any sensorimotor information during the trial is omitted, this results in

$$\boldsymbol{e}_j^{\mathrm{RW}} = 2e_j^{\gamma}(k_{\mathrm{end}}) \cdot d_j(k_{\mathrm{end}}). \tag{5.5}$$

In cases where the end-point error is not available and reward information is only binary, error for explicit reward learning is calculated as

$$\boldsymbol{e}_j^{\mathrm{RW}} = \begin{cases} 0 \; if \; \boldsymbol{e}_j^{\mathrm{RW}} \leq r \\ 1 \; if \; \boldsymbol{e}_j^{\mathrm{RW}} > r \end{cases}, \tag{5.6}$$

where r is the radius of the target.

To model the human learning process, we used an iterative learning control (ILC) system, where a current-iteration feedback controller is incorporated with ILC in the parallel architecture and combined with learning based on reward prediction error as seen on Figure 5.1. We modelled the human learning process by combining the input from the explicit reward learning $\boldsymbol{u}_j^{\mathrm{RW}}(k)$, input from the feedforward sensorimotor adaptation $\boldsymbol{u}_j^{\mathrm{SM}}(k)$ and input from the feedback control $\boldsymbol{u}_j^{\mathrm{FB}}(k)$, into the motor command

$$\boldsymbol{u}_j(k) = \boldsymbol{u}_j^{\mathrm{RW}}(k) + \boldsymbol{u}_j^{\mathrm{SM}}(k) + \boldsymbol{u}_j^{\mathrm{FB}}(k). \tag{5.7}$$

The model assumes an additive effect of sensorimotor adaptation and explicit reward learning during a single trial, however, the amount of explicit reward learning has an effect on the sensorimotor learning in the succeeding trials.

Learning based on reward prediction error is used to plan the direction of movement. It either takes into account the difference between the actual and the optimal movement of the cursor in the previous simulation trial or explores different movements to determine the movement in the next simulation trial as

$$\boldsymbol{u}_j^{\mathrm{RW}}(k) = \boldsymbol{u}_0(k) + \boldsymbol{u}_{j-1}^{\mathrm{RW}}(k) + p \, \boldsymbol{e}_{j-1}^{\mathrm{RW}}, \tag{5.8}$$

where p is an explicit reward learning parameter. In the case of end-point error learning, parameter p is set to a fixed value, while during the exploration phase in the binary reward learning parameter p is sampled from a uniform distribution. When a successful movement is found, motor commands for this movement are stored in the memory module. $\boldsymbol{u}_0$ represents the previous motor knowledge for an unperturbed movement that produces a straight minimum jerk trajectory (Flash & Hogan, 1985) between the initial point and target. This knowledge is stored in the memory / fallback module and represents the initial conditions of motor commands for the first simulation trial ($j = 1$). Furthermore, these

initial conditions are used when the perturbation is removed and the system is aware of the lack of perturbation (de-adaptation trials). It is not used during catch trials, since the system doesn't know the perturbation is removed.

Learning based on sensory prediction error is used for feedforward sensorimotor adaptation that is updated from one simulation trial to the next as

$$\boldsymbol{u}_j^{\mathrm{SM}}(k) = b(\boldsymbol{u}_{j-1}^{\mathrm{SM}}(k) + \mathbf{L}\,\boldsymbol{e}_{j-1}(k)) + \boldsymbol{n}_{\mathrm{m}} \tag{5.9}$$

where b is forgetting factor set to 0.995, $\mathbf{L}$ is a matrix representing the feedforward sensorimotor adaptation parameters and $\boldsymbol{n}_{\mathrm{m}}$ is normally distributed motor noise with variance $\sigma_{fd}^2 = 0.02$ for distance and $\sigma_{f\gamma}^2 = 0.8$ for angle.

Input from the feedback control $\boldsymbol{u}_j^{\mathrm{FB}}(k)$ is updated from one time-step to the next during a single simulation trial as

$$\boldsymbol{u}_j^{\mathrm{FB}}(k) = \mathbf{C}\,\boldsymbol{e}_j(k-1), \tag{5.10}$$

where $\mathbf{C}$ is a matrix representing feedback sensorimotor parameters. We incorporated a delayed reaction time of 0.3 s for feedback control, based on the experimental data.

5.2.1.1 Simulation Setup

The described model was used to carry out simulations of the reaching task described in experiment section (2.4.2 Experiment). There were 40 time steps of 0.02 s each. To allow for a direct comparison between the experimental and simulation results, the number of simulation trials was set to 80 and the presence of visuomotor rotation was included in the same number of simulation trials as in the experiment. Simulations included the baseline trials, trials with perturbation in one direction, 2 catch trials and de-adaptation trials (as shown in the top part of Figure 2.1). The system had a visuomotor rotation feature, where the angle of rotation was determined by the angle β. During the simulation, initial trials had β set to 0. For the following trials, the angle β was set to 30°, except for the two catch trials and de-adaptation trials, where β was again set to 0. To obtain representative values, we ran the simulations for each combination of parameter values 50 times and then calculated the average values. Sensorimotor adaptation and reward learning were first simulated separately. Sensorimotor adaptation was based on sensory prediction error throughout the movement and explicit reward learning was based on reward prediction error at the end of the movement. Here we distinguished between the reward in the form of end-point error (as in the group EPE) and a binary reward (as in the group RWD). Sensorimotor adaptation regarded all the sensory information during the trial and included feedback control, while explicit reward learning disregarded sensory information during the trial and did not include feedback control. Furthermore, the sensorimotor adaptation and explicit reward learning processes were then combined to simulate the learning observed in experimental study. Sensorimotor adaptation was combined with different rates of explicit reward learning where end-point error was taken as a reward prediction error. Following, the end-point error learning without feedback control was combined with different rates of sensorimotor adaptation.

Parameter p and parameters in matrices $\mathbf{L}$ and $\mathbf{C}$ model the learning rates and were changed for different simulation runs. The presented results include a range of different values between maximum and minimum parameter values where the model was stable. p models the learning rate of the explicit reward learning in the case of end-point error learning and also models the exploration in the case of binary reward learning, while $\mathbf{L}$ models the learning rate of the sensorimotor adaptation. To investigate the sensorimotor adaptation process, we carried out simulation runs where we modulated the parameters in matrix $\mathbf{L}$ while p was set to 0. To investigate the end-point error learning process, we carried out simulation runs where we modulated the parameter p while $\mathbf{L}$ was set to

$$\mathbf{L} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

To investigate binary reward learning process, we carried out simulation runs where parameter p was sampled from uniform distribution while $\mathbf{L}$ was again set to

$$\mathbf{L} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

$\mathbf{C}$ represents the PD feedback controller and was set to

$$\mathbf{C} = \begin{bmatrix} 0.4 & 5 & 0 & 0 \\ 0 & 0 & 0.4 & 5 \end{bmatrix}$$

for all simulation runs. Parameter $\mathbf{C}$ was chosen so that the feedback contribution best matches the experimental results. However, when we simulated trials where no sensory information was provided, there was no contribution of the feedback control due to the absence of sensory prediction error.

Additionally, we carried out simulation runs with a combination of both reward and sensorimotor adaptation, to capture the learning processes observed in the experimental study. Since the goal here was to simulate the experimental results, the parameter values used for both sensorimotor adaptation and explicit reward learning were chosen so that the simulation results best match the result from the experiment. Simulation capturing results from ERR group had parameters set to

$$\mathbf{L} = \begin{bmatrix} 0.02 & 0.9 & 0 & 0 \\ 0 & 0 & 0.02 & 0.9 \end{bmatrix}, p = 0.36.$$

Simulation capturing results from EPE group had parameters set to

$$\mathbf{L} = \begin{bmatrix} 0.0012 & 0.96 & 0 & 0 \\ 0 & 0 & 0.0012 & 0.096 \end{bmatrix}, p = 0.5.$$

5.2.1.2 Simulation Results

We first simulated sensorimotor adaptation based on sensory prediction error (Figure 5.2A-B), without taking into account the reward prediction error. This was done by setting the parameter p, which determines explicit reward learning, to zero and modulating parameters

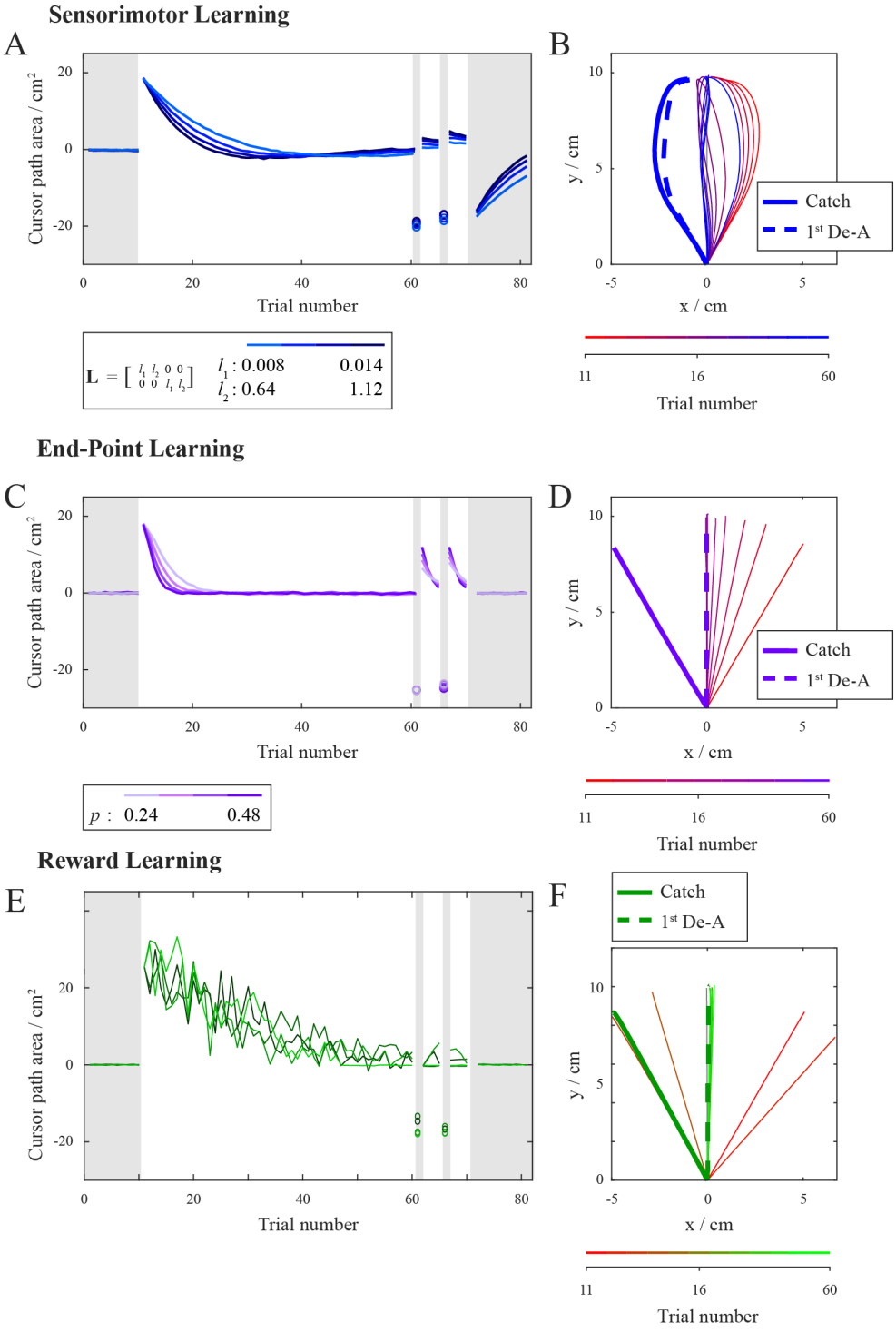


Figure 5.2: Simulation of different learning processes. (A) Mean values of Cursor path area showing the progression of sensorimotor adaptation. Different shades of blue represent different sensorimotor adaptation parameter values in matrix **L**. Shaded areas represent trials without perturbation and non-shaded areas represent trials with perturbation. (B) Example of simulated cursor trajectories for sensorimotor adaptation showing the initial four perturbed trials followed by every sixth trial, catch trial and 1st de-adaptation trial. The progression of the trials is represented with the transition of the red into blue colour.

Bold blue line depicts catch trial and bold dashed blue line depicts the 1st de-adaptation trial. (C) Mean values of Cursor path area showing the progression of end-point error learning. Different shades of violet represent different values of p for explicit reward learning. Shaded areas represent trials without perturbation and non-shaded areas represent trials with perturbation. (D) Example of simulated cursor trajectories for end-point error learning showing initial four perturbed trials followed by every sixth trial, catch trial and 1st de-adaptation trial. The progression of the trials is represented with the transition of the red into violet colour. Bold violet line depicts the catch trial and bold dashed violet line depicts the 1st de-adaptation trial. (E) Mean values of Cursor path area showing the progression of binary reward learning. Different shades of green represent different exploration runs. Shaded areas represent trials without perturbation and non-shaded areas represent trials with perturbation. (F) Example of simulated cursor trajectories for binary reward learning showing initial four perturbed trials followed by every sixth trial, catch trial and 1st de-adaptation trial. The progression of the trials is represented with the transition of the red into green colour. Bold green line depicts the catch trial and bold dotted green line depicts the 1st de-adaptation trial.

in the matrix **L**, which determines sensorimotor adaptation. Trajectories were updated from one simulation trial to the next based on the errors in position and velocity at every time-step. We also included input from feedback control which was updated from one time-step to the next during a single simulation trial. Initial trials without the perturbation resulted in straight cursor trajectories, since the model has previous motor knowledge for an unperturbed movement stored in the memory / fallback module. When β was set to 30° the cursor trajectory shifted in the right direction at the beginning of the trial and then moving closer to the middle as a result of the feedback controller. After a few trials, the model learned to adapt to the perturbation and Cursor path area decreased over trials. The slope of the decline depended on the different values of the parameters in **L** that affect the learning rate. During the catch trials, where β was set to 0, the same motor commands were used as in the previous trials with perturbation. This resulted in large aftereffect (bold blue trajectory on Figure 5.2B). During the de-adaptation trials (dotted blue line on Figure 5.2B), β was again set to 0 which resulted in similar aftereffect as observed during catch trials.

We then modelled explicit reward learning based on end-point error (Figure 5.2C-D) without sensorimotor adaptation and feedback control. Learning based on reward prediction error was used to plan the direction of movement. The sensory information during the trial was disregarded and only the information on the error at the final time step of the previous simulation trial was considered. This was done by setting all the parameters in matrix **L**, which determines sensorimotor adaptation, to zero and modulating the parameter p, which determines explicit reward learning. As in the case of sensorimotor adaptation, initial trials without the perturbation resulted in straight cursor trajectories. When β was set to 30°, the cursor trajectory shifted in the right direction without any corrections during the trial, resulting in a straight but angled cursor trajectory. After a few trials, the model learned to adapt to the perturbation and Cursor path area decreases over trials. The slope of the decline depends on the different values for parameter p that affect

the learning rate. Based on the end-point error, the model adjusted the direction of movement and planned the trajectory accordingly. During catch trials, where β was set to 0, the same motor commands were used as in the previous trials with perturbation. During catch trials, the planned trajectory was based on the previously learned perturbation and therefore resulted in significant aftereffects (bold violet trajectory on Figure 5.2D). On the other hand, in the de-adaptation trials, the new motion was based on the motor commands for the straight movement that were stored in the fallback module. Since there were no changes due to the sensorimotor adaptation and the changes resulting from explicit reward learning were disregarded due to the new information on the perturbation, the direction of movement was adjusted accordingly. During the de-adaptation trials (dashed violet line on Figure 5.2D), β was again set to 0 which resulted in a straight trajectory similar to the initial trials without perturbation.

Following, we modelled explicit reward learning based on binary error (Figure 5.2E-F) without sensorimotor adaptation and feedback control. Similarly, as in end-point learning, there was no sensory information provided during the simulated trials resulting in no contribution from the feedback controller. Additionally, there was no information about the end-point error. The only information provided was, if the trial was a success or a failure. If the trial was a success, motor commands were stored and used in the next trial. If the trial was a failure, we used exploration to determine the movement in the next simulation trial. In this case, parameter p was randomly determined from a uniform distribution to allow for exploration. Initial trials without the perturbation resulted in straight cursor trajectories. When β was set to 30° the cursor trajectory shifted in the right direction without any corrections during the trial, resulting in a straight but angled cursor trajectory, similar as in end-point learning. This was followed by multiple exploration trials before the model found a solution. After initial period of exploration, a successful solution is found and Cursor path area decreases over trials. Similar as in end-point error learning, the planned trajectory during catch trials was based on the previously learned perturbation and therefore resulted in significant aftereffects (bold green trajectory on Figure 5.2F). De-adaptation trials (dotted green line on Figure 5.2F) resulted in a straight trajectory similar to the initial trials without perturbation.

We then proceeded to simulate the results obtained in the experimental study, by combining the explicit reward learning and sensorimotor adaptation. Two different simulations were done, one with the sensory information provided throughout the trials as in the ERR group and the other without the sensory information and only with the end-point error information provided at the end of the trials as in the EPE group. We first modelled learning with sensory information provided (Figure 5.3A). Trials during perturbation and catch trials were similar as in previous simulation. However, Cursor path area values for de-adaptation trials were smaller than in previous simulation, indicating smaller aftereffects when the absence of perturbation was known. The aftereffects were influenced by the amount of explicit reward learning present. Without explicit reward learning the aftereffects were similar to the ones observed during the catch trials, while with higher learning rates of explicit reward learning the aftereffects decreased.

We then modelled learning without sensory information provided (Figure 5.3B). Trials during perturbation and catch trials were similar as in previous simulation. However,

Cursor path area values for the de-adaptation trials were higher as in the previous simulation, indicating higher aftereffects when the absence of perturbation was known. The aftereffects were influenced by the amount of sensorimotor adaptation present. Without the sensorimotor adaptation the aftereffects were not present, while with the higher learning rates of sensorimotor adaptation the aftereffects increased. Figure 5.3C-D shows example trajectories of catch and de-adaptation trials and average Cursor path area for baseline, catch and de-adaptation trials.

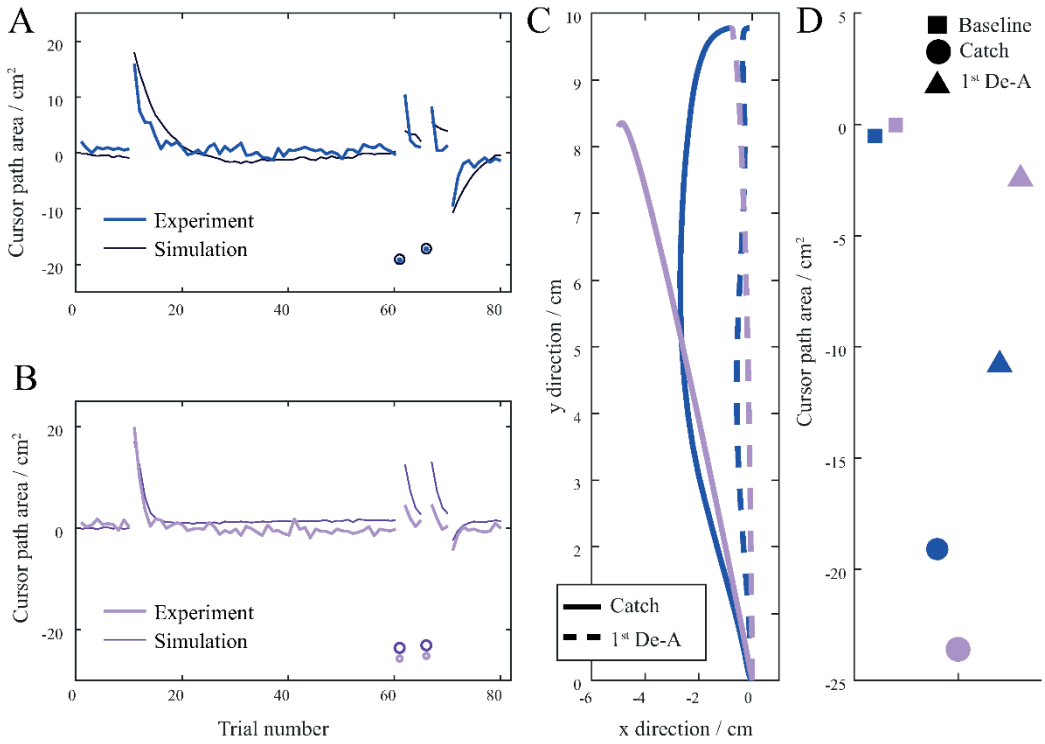


Figure 5.3: Simulation of a combination of sensorimotor and explicit reward learning. (A) Mean values of Cursor path area showing the progression of learning simulating results from the ERR group. Black represents simulation results and blue represents experimental results. (B) Mean values of Cursor path area showing the progression of learning simulating results from the EPE group. Dark violet represents simulation results and light violet represents experimental results. (C) Example of simulated trajectories showing the catch trials and 1st de-adaptation trials. Because there is no perturbation for these trials, cursor trajectories correspond to hand movement trajectories. Bold blue line depicts catch trial and bold dashed blue line depicts 1st de-adaptation trial for ERR group. Bold violet line depicts catch trial and bold dashed violet line depicts 1st de-adaptation trial for RWD group. (D) Average Cursor path area for baseline, catch and de-adaptation trials.

5.2.2 Adaptive ILC Model for Visuomotor Rotation

To better capture the mechanisms underlying visuomotor adaptation, we further refined the initial ILC-based modelling approach. Although the original implementation was able to reproduce most behavioural patterns observed in the visuomotor rotation experiment, it did not provide an optimal account of the adaptation process. In the standard ILC formulation, error is accumulated across time steps and used to update motor commands

directly. Because early modifications affect the subsequent evolution of the trajectory, corrections that reduce error at initial time steps propagate forward and may produce overcompensation in later phases of the movement. As a result, the model lacks a mechanism to regulate how corrections are distributed over time, leading to excessive compensation toward the end of the movement. To address these limitations, an alternative adaptive control model was developed, based on the same principles as the original ILC framework but incorporating controller-level adaptations rather than direct command updates (Figure 5.4).

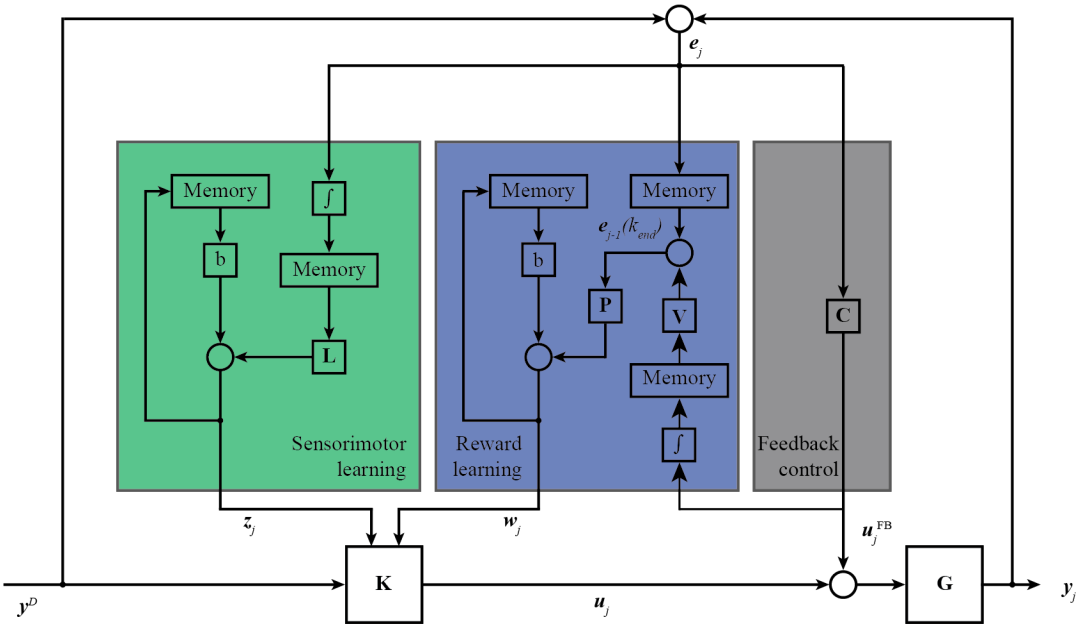


Figure 5.4: Architecture of the adaptive computational model. Controller K computes control signals based on sensorimotor learning parameter $\mathbf{z}$ and reward learning parameter $\mathbf{w}$. Control signal $\boldsymbol{u}_j$ is combined with feedback contribution $\boldsymbol{u}_j^{\mathrm{FB}}$ and motor commands are generated to the plant G. The integrator block $\int$ represents the calculation of the error for sensorimotor learning and additional feedback factor for reward learning. Reward error can be end-point error or binary information. Memory module represents motor commands from the previous simulation trial.

As with the previous ILC model, the system was a discrete-time, linear time-invariant MIMO system:

$$\boldsymbol{x}_j(k+1) = \mathbf{A}\boldsymbol{x}_j(k) + \mathbf{B}(\boldsymbol{u}_j(k) + \boldsymbol{u}_j^{FB}(k)), \tag{5.11}$$

with $\boldsymbol{u}_j^{FB}$ calculated as in Eq. (5.10) and output $\boldsymbol{y}_j$ as in Eq. (5.2). It incorporated explicit reward learning, feedforward sensorimotor adaptation, and feedback control.

As opposed to the previous model where motor knowledge for an unperturbed movement that produces a straight minimum-jerk trajectory is stored, here, a straight minimum jerk trajectory is calculated for every movement based on the equations from (Flash & Hogan, 1985). Control force in the radial direction is calculated as:

$$f_j^d(k) = \frac{\left(d_j^D(k_{\text{end}}) + z_j^d + w_j^d\right)}{T^2} * (120{\tau_k}^3 - 180{\tau_k}^2 + 60\tau_k) * m \tag{5.12}$$

where $\tau_k = k/k_{\text{end}}$ is normalized time, $d_j^D(k_{\text{end}})$ is desired final radial distance, z_j^d and w_j^d are added learning components defined below, and T is total time of movement. Since we are assuming straight movement trajectory, the control force in the polar angle direction $f_j^\gamma(k) = 0$. With learning, we are changing the initial polar angle as

$$\gamma_j(k_\circ) = \gamma_j^D(k_{end}) + z_j^\gamma + w_j^\gamma \tag{5.13}$$

where $\gamma_j^D(k_{end})$ is the desired final polar angle, z_j^γ and w_j^γ are added learning components. To adapt to the visuomotor perturbation, the learning components **z** and **w** are updated based on the error from the previous iteration, calculated as in Eq. (5.3).

During the sensorimotor learning, learning component **z** is changed as

$$\mathbf{z}_j = \begin{bmatrix} z_j^d \\ z_j^\gamma \end{bmatrix} = \left(\mathbf{z}_{j-1} * b^{SM}\right) - \mathbf{L} \sum_{l=1}^{k_{\text{end}}} \boldsymbol{e}_{j-1}(l) \tag{5.14}$$

while during reward learning, learning component **w** is changed as

$$\mathbf{w}_j = \begin{bmatrix} w_j^d \\ w_j^\gamma \end{bmatrix} = \left(\mathbf{w}_{j-1} * b^{RW}\right) - \mathbf{P}\, \boldsymbol{e}_{j-1}^{RW} \tag{5.15}$$

where **L** is feedforward sensorimotor adaptation parameter and **P** is an explicit reward learning parameter. Forgetting factors were set to $b^{SM} = 0.995$ and $b^{RW} = 0.99$.

Reward error, with an addition of the added feedback factor contributing to the overall reward error, was calculated as

$$\boldsymbol{e}_j^{\text{RW}} = \begin{bmatrix} e_j^d(k_{\text{end}}) \\ e_j^\gamma(k_{\text{end}}) \end{bmatrix} + \mathbf{V} \sum_{l=1}^{k_{\text{end}}} u_j^{FB}(l). \tag{5.16}$$

In cases where reward information is only binary, the reward learning component is updated based on task success. If the final cursor position reaches the target region, the previous angular strategy is retained. If the target is missed, a new angular strategy is sampled stochastically:

$$w_j^\gamma = \begin{cases} w_{j-1}^\gamma, & \text{if } \left\|e_j(k_{\text{end}})\right\| \le r, \\ \epsilon_j^\gamma, & \text{otherwise,} \end{cases} \tag{5.17}$$

where r is the target radius and $\epsilon_j^\gamma \sim N(0, \sigma^2)$ is a randomly sampled angular correction.

5.2.2.1 Simulation Setup

The described adaptive ILC model was used to carry out simulations of the reaching task with visuomotor rotation described in the experiment section (3.4.2 Experiment). The simulation mimicked the experimental design from the visuomotor rotation study. There were 40 time steps of 0.02s each and a visuomotor rotation feature, where the angle of

rotation was determined by the angle β. To obtain representative values, we ran the simulations for each combination of parameter values 8 times and then calculated the average values.

Parameters **L** and **P** were changed for different simulation runs. In the case of end-point error learning we modulated the parameter **P** and in the case of binary reward learning, learning component **w** was sampled from normal distribution. In both cases the parameter **L** was set to

$$\mathbf{L} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

On the other hand, to investigate the sensorimotor adaptation process, we carried out simulation runs where we modulated the parameters in matrix **L** while **P** was set to

$$\mathbf{P} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

C represents the PD feedback controller and was set to

$$\mathbf{C} = \begin{bmatrix} 0.0095 & 0.003 & 0 & 0 \\ 0 & 0 & 0.0095 & 0.003 \end{bmatrix}$$

for all simulation runs where sensory information was provided. Parameter **C** was chosen so that the feedback contribution best matches the experimental results.

Additionally, we carried out simulation runs with a combination of both reward and sensorimotor adaptation, to capture the learning processes observed in the experimental study. Here, the parameter values were chosen so that the model best captures the experimental results. Three different simulations were done, one with the sensory information provided throughout the trials as in the ERR group, one without the sensory information and only with the end-point error information provided at the end of the trials as in the EPE group and one without the sensory information and only with the binary information on the success of the trial as in RWD group. All simulations included both sensorimotor adaptation and reward learning, with differing levels of each. Simulation capturing results from ERR group had parameters set to

$$\mathbf{L} = \begin{bmatrix} 0.009 & 0.0009 & 0 & 0 \\ 0 & 0 & 0.009 & 0.0009 \end{bmatrix}, \mathbf{P} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \end{bmatrix}.$$

Simulation capturing results from EPE group had parameters set to

$$\mathbf{L} = \begin{bmatrix} 0.0015 & 0.00015 & 0 & 0 \\ 0 & 0 & 0.0015 & 0.00015 \end{bmatrix}, \mathbf{P} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix},$$

allowing for higher contribution of reward learning and smaller contribution of sensorimotor adaptation compared to the simulation capturing results from ERR group. Simulation capturing results from RWD group was sampling values for learning component **w** from a normal distribution around the perceived target area until the correct solution was found. Parameters were set to

$$\mathbf{L} = \begin{bmatrix} 0.001 & 0.0001 & 0 & 0 \\ 0 & 0 & 0.001 & 0.0001 \end{bmatrix}, \mathbf{P} = \begin{bmatrix} 0.07 & 0 \\ 0 & 0.07 \end{bmatrix}$$

to allow for a small contribution of sensorimotor adaptation and end-point error learning.

5.2.2.2 Simulation Results

The model was able to capture the learning behaviour observed in experimental results and replicate the differences between groups. Results of the simulation are shown in Figure 5.5. A straight minimum jerk trajectory was calculated for every movement resulting in straight cursor trajectories during the initial trials without the perturbation. When perturbation started, cursor trajectories shifted in the right direction. After a few trials, the model learned to adapt to the perturbation and Cursor path area decreased over trials. All learning processes exhibited aftereffects during catch trials, while aftereffects during de-adaptation trials were only present during sensorimotor adaptation. Overall, the adaptive ILC model displays similar results to the ILC model, while being more stable and allowing for wider options in parameter values.

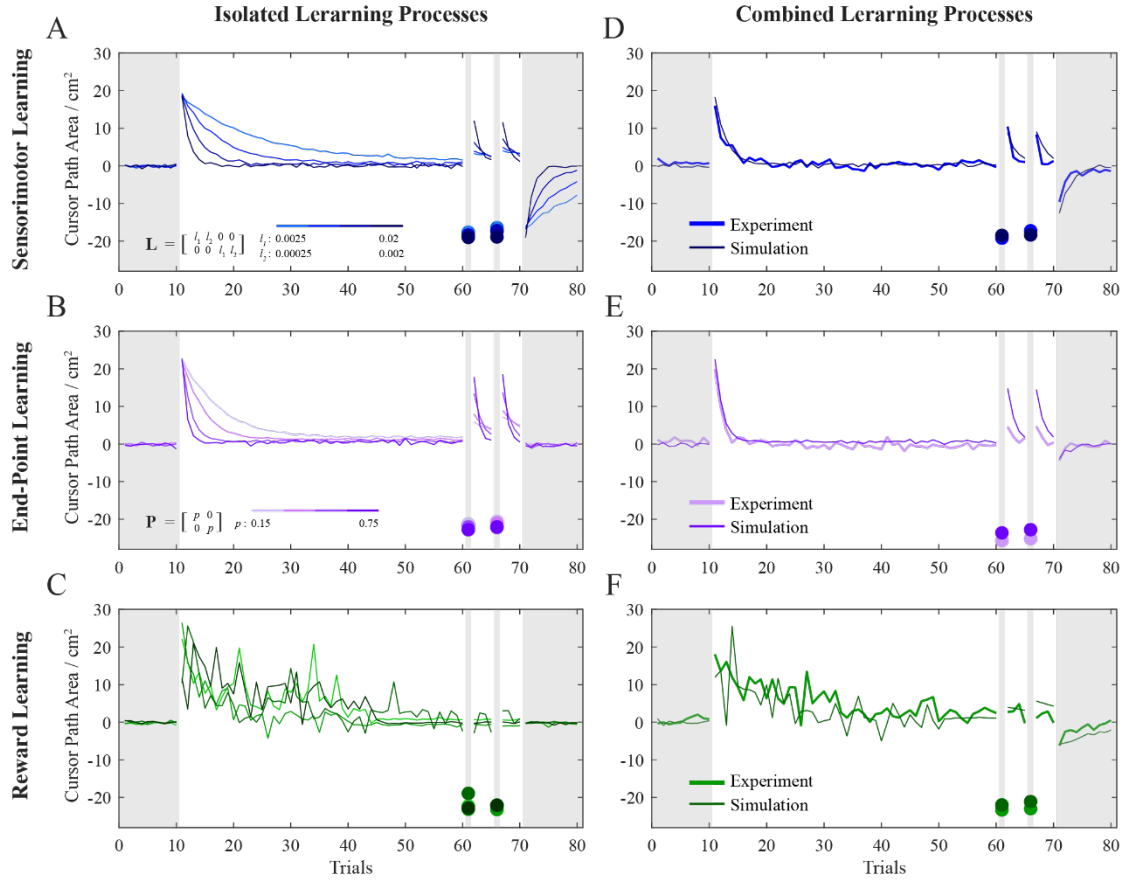


Figure 5.5: Simulation results of adaptive ILC model for visuomotor rotation. (A-C) Mean values of Cursor path area showing the progression of sensorimotor adaptation, end-point error learning and explicit reward learning, respectively. Different shades of blue represent different sensorimotor adaptation parameter values in matrix **L**, different shades of violet represent different values in matrix **P** for explicit reward learning and different

shades of green represent different exploration runs. (D-F) Mean values of Cursor path area showing the progression of learning simulating results from the ERR, EPE and RWD groups. Dark colours represent simulation results and lighter colours represent experimental results. Shaded areas represent trials without perturbation and non-shaded areas represent trials with perturbation.

5.2.3 Adaptive ILC Model for Force-Field Perturbation

Building on the adaptive ILC model developed for visuomotor perturbations, the same architectural framework was applied to model force-field adaptation. While the core structure of the model remained unchanged, modifications were introduced to account for the dynamic nature of the force-field perturbation. Specifically, the coordinate system was changed from polar to cartesian, with coordinates X and Y:

$$\boldsymbol{x}_j(k+1) = \mathbf{A}\boldsymbol{x}_j(k) + \mathbf{B}(\boldsymbol{u}_j(k) + \boldsymbol{u}_j^{FB}(k) + \boldsymbol{u}_j^{P}(k)) \tag{5.18}$$

$$\boldsymbol{y}_j(k) = \boldsymbol{x}_j(k), \tag{5.19}$$

with input from feedback control $\boldsymbol{u}_j^{FB}$ calculated as in Eq. (5.10) and force field perturbation $\boldsymbol{u}_j^{P}$ determined based on Eq. (3.1).

Control vector was defined as

$$\boldsymbol{u}_j(k) = \begin{bmatrix} f_j^X(k) \\ f_j^Y(k) \end{bmatrix},$$

state vector as

$$\boldsymbol{x}_j(k) = \begin{bmatrix} X_j(k) \\ \dot{X}_j(k) \\ \mathrm{Y}_j(\mathrm{k}) \\ \dot{Y}_j(k) \end{bmatrix},$$

input matrix as

$$\mathbf{B} = \begin{bmatrix} 0 & 0 \\ 1/m & 0 \\ 0 & 0 \\ 0 & 1/m \end{bmatrix}$$

and error vector as

$$\boldsymbol{e}_j(k) = \begin{bmatrix} e_j^X(k) \\ e_j^{\dot{X}}(k) \\ e_j^Y(\mathrm{k}) \\ e_j^{\dot{Y}}(k) \end{bmatrix}.$$

A straight minimum jerk trajectory was calculated for every movement based on the equations from Flash & Hogan, 1985 with an added factor allowing for force field adaptation:

$$\begin{bmatrix} f_j^X(k) \\ f_j^Y(k) \end{bmatrix} = \begin{bmatrix} X_j(k_{end}) \\ Y_j(k_{end}) \end{bmatrix} * ((120k^3 - 180k^2 + 60k) + (\mathbf{z}_j + \mathbf{w}_j) * (30{\tau_k}^4 - 60{\tau_k}^3 + 30{\tau_k}^2)) * m/T^2 \tag{5.20}$$

Where $\tau_k = k/k_{\text{end}}$ is normalized time, z_j^d and w_j^d are added learning components defined below, and T is total time of movement. $\mathbf{z}_j$ represents a sensorimotor learning component and $\mathbf{w}_j$ represents a reward learning component.

During the sensorimotor learning, learning component $\mathbf{z}$ is changed as

$$\mathbf{z}_j = \begin{bmatrix} z_j^X \\ z_j^Y \end{bmatrix} = \left(\mathbf{z}_{j-1} * b^{SM}\right) - \mathbf{L} \sum_{l=1}^{k_{\text{end}}} \boldsymbol{e}_{j-1}(l) \tag{5.21}$$

while during reward learning, learning component $\mathbf{w}$ is changed as

$$\mathbf{w}_j = \begin{bmatrix} w_j^X \\ w_j^Y \end{bmatrix} = b^{RW}\mathbf{w}_{j-1} - \mathbf{P}\,\boldsymbol{e}_{j-1}^{RW} \tag{5.22}$$

where $\mathbf{L}$ is feedforward sensorimotor adaptation parameter and $\mathbf{P}$ is an explicit reward learning parameter. Forgetting factors were set to $b^{SM} = 0.995$ and $b^{RW} = 0.99$.

Reward error, with an addition of the added feedback factor contributing to the overall reward error, was calculated as

$$\boldsymbol{e}_j^{\text{RW}} = \begin{bmatrix} e_j^X(k_{\text{end}}) \\ e_j^Y(k_{\text{end}}) \end{bmatrix} + \mathbf{V} \sum_{l=1}^{k_{\text{end}}} u_j^{FB}(l). \tag{5.23}$$

In cases where reward information is only binary, the reward learning component is updated based on task success. If the final cursor position reaches the target region, the previous horizontal strategy is retained. Otherwise, a new horizontal strategy is sampled stochastically:

$$w_j^X = \begin{cases} w_{j-1}^X, & \text{if } \left\|e_j(k_{\text{end}})\right\| \le r, \\ \epsilon_j^X, & \text{otherwise,} \end{cases} \tag{5.24}$$

where r is the target radius and $\epsilon_j^X \sim N(-\alpha\, sign(\kappa), \sigma^2)$ is horizontal correction biased opposite to the perturbation direction.

5.2.3.1 Simulation Setup

The described adaptive ILC model was used to simulate the reaching task with force-field perturbation described in the experiment section (Experiment). The simulation mimicked the experimental design of the force-field study and consisted of 300 trials including baseline, perturbation, catch and de-adaptation trials. Each trial contained 40 time steps of 0.02s. During perturbation trials, there was force-field perturbation included, determined based on Eq. (3.1). As in the previous simulations, individual learning processes then combined to reproduce the adaptation observed in the experimental study. To obtain

representative results, each combination of parameter values was simulated 10 times, and average performance measures were computed.

Parameters **L** and **P** were varied across simulation runs. In the case of end-point error learning we modulated the parameter **P** and in the case of binary reward learning parameter **P** was sampled from normal distribution. In both cases the parameter **L** was set to

$$\mathbf{L} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

On the other hand, to investigate the sensorimotor adaptation process, we carried out simulation runs where we modulated the parameters in matrix **L** while **P** was set to

$$\mathbf{P} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

C represents the PD feedback controller and was set to

$$\mathbf{C} = \begin{bmatrix} 13.5 & 4.5 & 0 & 0 \\ 0 & 0 & 13.5 & 4.5 \end{bmatrix}$$

for all simulation runs where sensory information was provided. Parameter **C** was chosen so that the feedback contribution best matches the experimental results.

Finally, simulations combining both reward-based and sensorimotor adaptation were performed to reproduce the behavioural patterns observed experimentally. Parameter values were chosen to best capture the experimental results. Three simulation conditions were implemented to represent the different experimental groups. Simulation capturing results from ERR group had parameters set to

$$\mathbf{L} = \begin{bmatrix} 0.07 & 0.007 & 0 & 0 \\ 0 & 0 & 0.07 & 0.007 \end{bmatrix}, \mathbf{P} = \begin{bmatrix} 0.0025 & 0 \\ 0 & 0.0025 \end{bmatrix}.$$

Simulation capturing results from EPE group had parameters set to

$$\mathbf{L} = \begin{bmatrix} 0.07 & 0.007 & 0 & 0 \\ 0 & 0 & 0.07 & 0.007 \end{bmatrix}, \mathbf{P} = \begin{bmatrix} 0.05 & 0 \\ 0 & 0.05 \end{bmatrix}.$$

Simulation capturing results from RWD group was sampling values for learning component **w** from a normal distribution around the perceived target area until the correct solution was found. Parameters were set to

$$\mathbf{L} = \begin{bmatrix} 0.05 & 0.005 & 0 & 0 \\ 0 & 0 & 0.05 & 0.005 \end{bmatrix}, \mathbf{P} = \begin{bmatrix} 0.035 & 0 \\ 0 & 0.035 \end{bmatrix}.$$

All simulations included the sensory information during the trial and had the contribution of feedback control, since the proprioceptive information was available for all groups.

5.2.3.2 Simulation Results

The model was able to capture the learning behaviour observed in experimental results and replicate the differences between groups. Results of the simulation are shown in Figure 5.6. A straight minimum jerk trajectory was calculated for every movement. During the

initial trials without the perturbation, movement trajectories were straight for both sensorimotor and reward learning. When perturbation started, perturbation force was forcing movements away from the target. Sensorimotor adaptation was simulated by modulating parameters in the matrix **L**, end-point error learning was simulated by modulating parameters in the matrix **P** and binary reward learning was simulated by sampling learning component **w** from a normal distribution around the perceived target area. As in the previous simulations, only the sensorimotor adaptation included feedback control, while explicit reward learning disregarded sensory information during the trial. We observed comparable results to the previous simulations where after a few trials, the model learned to adapt to the perturbation and kinematic error decreased over trials, all learning processes exhibited aftereffects during catch trials while aftereffects during de-adaptation trials were only present during sensorimotor adaptation.

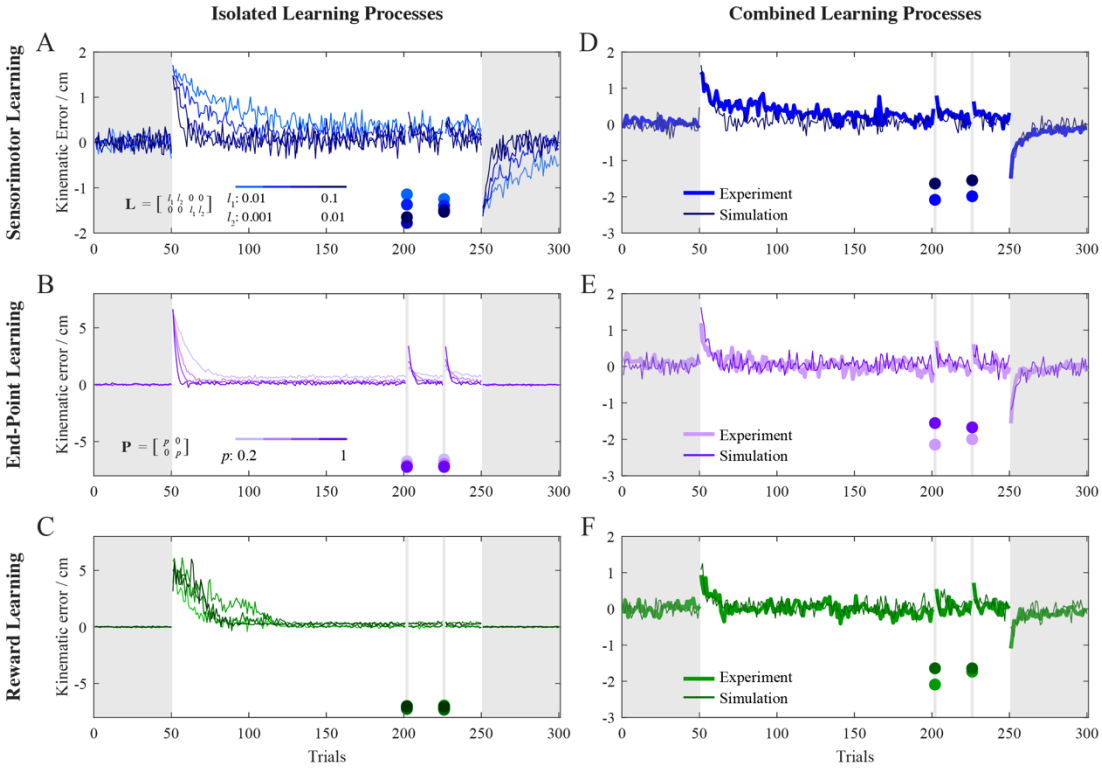


Figure 5.6: Simulation results of adaptive ILC model for force-field perturbation. (A-C) Mean values of Kinematic error showing the progression of sensorimotor adaptation, end-point error learning and explicit reward learning, respectively. Different shades of blue represent different sensorimotor adaptation parameter values in matrix **L**, different shades of violet represent different values in matrix **P** for explicit reward learning and different shades of green represent different exploration runs. (D-F) Mean values of Kinematic error showing the progression of learning simulating results from the ERR, EPE and RWD groups. Dark colours represent simulation results and lighter colours represent experimental results. Shaded areas represent trials without perturbation and non-shaded areas represent trials with perturbation.

5.3 Discussion

The sensorimotor adaptation and explicit reward learning processes cannot be separately observed in human motor learning due to the inability to perfectly separate them or inhibit one of them. To investigate the individual contributions of either sensorimotor adaptation or explicit reward learning on its own, we simulated different learning processes separately in a computational model. During simulation of sensorimotor adaptation, internal model adapts to the new perturbation, resulting in new motor commands. The changes to the motor commands are on the level of an internal model and can be looked at as changed motor commands for a movement resulting in a straight cursor trajectory. The same commands are then used in the case when perturbation is no longer present. Even if the model considers the lack of perturbation, the motor commands do not change since the model treats them as correct for a straight movement. Differently, explicit reward learning (either based on end-point error or binary reward) does not affect the internal model and does not change the motor commands directly. By using the reward error for the whole movement in the next trial, it in effect changes the direction of movement, or rather plans the movement toward a different target in the case of visuomotor rotation. In the case of force-field perturbations, it instead generates compensatory forces that push the hand in the opposite direction of the perturbation. This coincides with experiments that show participants were aiming in the different direction and knew their hand position was changed (Izawa & Shadmehr, 2011). Overall, the simulation results show how the learning proceeds regardless of the prediction error as well as it shows how the aftereffects are dependent on the sensory prediction error and do not occur during explicit reward learning.

Our experimental results clearly show two distinct processes of motor learning with different retention rates. It has been previously proposed that motor learning is comprised of two interacting processes, a fast process that adapts and decays quickly and a slower process that adapts and decays more gradually (Howard et al., 2015; Smith et al., 2006). Computationally, this was described by a dual-rate model that incorporates different timescales and retention rates (Smith et al., 2006). A dual-rate model does not address implicit and explicit learning directly; however, it has been suggested that the fast and slow processes of the dual-rate model may correspond to the explicit and implicit components of learning (Heald et al., 2021). A dual-rate model is a single context model and cannot explain the differences between the catch and the de-adaptation trials as well as the differences in switching between perturbations that were observed in our experiment (Heald et al., 2021). In contrast, our computational model accounts for the different timescales and retention rates with sensorimotor and reward learning and shows how multiple-contexts (visual cues) have different effects during sensorimotor or reward learning. It is also able to account for differences in catch and de-adaptation trials. Moreover, current models of motor learning usually rely on either sensory prediction error or reward prediction error (McDougle et al., 2016). However, recent behavioural experiments are showing that different learning processes respond to distinct error signals (Izawa & Shadmehr, 2011). Computational models that aim to represent human motor learning as a whole should take into account both sensorimotor adaptation based on sensory prediction error and explicit reward learning based on reward prediction error. The

advantage of our computational model is that it separately accounts for sensorimotor and reward prediction errors.

Even though the experiment could not completely separate explicit reward learning and sensorimotor adaptation, the simulation allowed us to look at both separately and compare them. We simulated motor learning with ILC where trials were updated from iteration to iteration. Even though we successfully simulated the majority of the trials, it ended up not being the optimal solution. In ILC, error is summed over time-steps so contribution of a basic feed-forward controller results in a curved cursor trajectory even when feedback contribution is omitted. Therefore, we designed an adaptive ILC model where changes are made to the controller and not directly to the motor commands as in ILC, the planned trajectory then also takes into account the contributions of the previous steps. This allows for a better adaptation in the case, where the change in the first step affects the later steps so that no additional change is necessary. Additionally, the schematic representation of the learning process with ILC is not the best representation of the learning processes taking place in the human brain, because it directly modifies the motor commands instead of the feedback motor commands, parameters of the feedforward internal model and the planned motion. A more sophisticated adaptive control model based on the same concept as the simple ILC model used to clarify sensorimotor and reward learning processes was able to deal with these. It is important for the future work to also include comparing the model to standard models of motor adaptation.

Chapter 6

Conclusions

This dissertation examined the interplay between sensorimotor and reward learning in human motor adaptation, revealing how these two processes contribute distinctly yet complementarily to adaptive motor control. Through different experimental studies we demonstrated that motor learning is not a unitary process, but a dynamic integration of multiple learning mechanisms governed by different error signals, learning timescales, and task demands. These findings confirm the hypothesis H1 that reward learning and sensorimotor learning represent two distinct processes of motor adaptation.

Our results further support the hypothesis H2 that sensorimotor adaptation is primarily an implicit process driven by sensory prediction error and responsible for modifying the internal model. Across both visuomotor and force-field perturbation paradigms, sensorimotor learning produced robust aftereffects, consistent with stable internal model updating and limited flexibility under randomly changing conditions.

In contrast, the findings confirm the hypothesis H3 that reward-based learning operates as an explicit process driven by reward consequences and does not alter the internal model. As predicted, reward learning did not produce aftereffects and enabled rapid adaptation in randomly changing environments through strategic adjustments.

The whole-body motor adaptation experiments provide support for the hypothesis H4 that more complex and ecologically relevant motor tasks require the combined engagement of reward-based and sensorimotor learning mechanisms. Extending our findings into ecologically valid contexts, we demonstrated that humans do not simply optimize performance under idealized conditions but adapt according to task-level outcomes related to movement efficiency and safety. Participants dynamically balanced movement optimization and injury avoidance, exhibiting a high-level ecological control strategy that shifted depending on success or failure. This behavioural switching is consistent with a win-stay/lose-shift policy and reflects the integration of reinforcement learning, internal model adaptation, and feedback control within a hierarchical control architecture.

Finally, the computational modelling results support the hypothesis H5 that an adaptive control architecture integrating sensorimotor learning, reward learning, and feedback control can reproduce the distinct behavioural patterns observed experimentally. The proposed computational framework successfully captured both the dissociation and interaction of these learning mechanisms across paradigms.

Together, these findings support a hierarchical view of motor learning in which high-level processes shaped by task success and ecological relevance guide reward-based planning, while lower-level processes refine execution through implicit adaptation and feedback control. This framework reconciles previously competing models by showing that sensorimotor and reward learning operate in parallel, often simultaneously, and are selectively engaged depending on the nature of the task, perturbation, and error feedback. This work advances our understanding of how humans achieve flexible, efficient, and robust motor behaviour in real-world environments. It emphasizes the need to consider both explicit and implicit learning processes in models of motor control and suggests that future interventions, training programs, and assistive technologies should adaptively engage both mechanisms based on context and individual needs.

References

Abe, M., Schambra, H., Wassermann, E. M., Luckenbaugh, D., Schweighofer, N., & Cohen, L. G. (2011). Reward Improves Long-Term Retention of a Motor Memory through Induction of Offline Memory Gains. *Current Biology*, *21*(7), 557–562. https://doi.org/10.1016/j.cub.2011.02.030

Akazawa, K., Milner, T. E., & Stein, R. B. (1983). Modulation of reflex EMG and stiffness in response to stretch of human finger muscle. *Journal of Neurophysiology*, *49*(1), 16–27. https://doi.org/10.1152/jn.1983.49.1.16

Albert, S. T., Jang, J., Modchalingam, S., 't Hart, B. M., Henriques, D., Lerner, G., Della-Maggiore, V., Haith, A. M., Krakauer, J. W., & Shadmehr, R. (2022). Competition between parallel sensorimotor learning systems. *ELife*, *11*. https://doi.org/10.7554/eLife.65361

Albert, S. T., Jang, J., Sheahan, H. R., Teunissen, L., Vandevoorde, K., Herzfeld, D. J., & Shadmehr, R. (2021). An implicit memory of errors limits human sensorimotor adaptation. *Nature Human Behaviour*, *5*(7), 920–934. https://doi.org/10.1038/s41562-020-01036-x

Averbeck, B. B. (2015). Theory of Choice in Bandit, Information Sampling and Foraging Tasks. *PLOS Computational Biology*, *11*(3), e1004164. https://doi.org/10.1371/journal.pcbi.1004164

Babič, J., Kunavar, T., Oztop, E., & Kawato, M. (2025). Success-efficient/failure-safe strategy for hierarchical reinforcement motor learning. *PLOS Computational Biology*, *21*(5), e1013089. https://doi.org/10.1371/journal.pcbi.1013089

Babič, J., Oztop, E., & Kawato, M. (2016). Human motor adaptation in whole body motion. *Scientific Reports*, *6*(1), 32868. https://doi.org/10.1038/srep32868

Balasubramanian, S., Melendez-Calderon, A., Roby-Brami, A., & Burdet, E. (2015). On the analysis of movement smoothness. *Journal of NeuroEngineering and Rehabilitation*, *12*(1), 112. https://doi.org/10.1186/s12984-015-0090-9

Bauby, C. E., & Kuo, A. D. (2000). Active control of lateral balance in human walking. *Journal of Biomechanics*, *33*(11), 1433–1440. https://doi.org/10.1016/S0021-9290(00)00101-9

Bond, K. M., & Taylor, J. A. (2015). Flexible explicit but rigid implicit learning in a visuomotor adaptation task. *Journal of Neurophysiology*, *113*(10), 3836–3849. https://doi.org/10.1152/jn.00009.2015

Braun, D. A., Nagengast, A. J., & Wolpert, D. M. (2011). Risk-Sensitivity in Sensorimotor Control. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00001

Burdet, E., Osu, R., Franklin, D. W., Milner, T. E., & Kawato, M. (2001). The central nervous system stabilizes unstable dynamics by learning optimal impedance. *Nature*, *414*(6862), 446–449. https://doi.org/10.1038/35106566

Bütefisch, C. M., Davis, B. C., Wise, S. P., Sawaki, L., Kopylev, L., Classen, J., & Cohen, L. G. (2000). Mechanisms of use-dependent plasticity in the human motor cortex. *Proceedings of the National Academy of Sciences*, *97*(7), 3661–3665. https://doi.org/10.1073/pnas.97.7.3661

Calalo, J. A., Roth, A. M., Lokesh, R., Sullivan, S. R., Wong, J. D., Semrau, J. A., & Cashaback, J. G. A. (2023). The sensorimotor system modulates muscular co-contraction relative to visuomotor feedback responses to regulate movement variability. *Journal of Neurophysiology*, *129*(4), 751–766. https://doi.org/10.1152/jn.00472.2022

Cashaback, J. G. A., McGregor, H. R., Mohatarem, A., & Gribble, P. L. (2017). Dissociating error-based and reinforcement-based loss functions during sensorimotor learning. *PLOS Computational Biology*, *13*(7), e1005623. https://doi.org/10.1371/journal.pcbi.1005623

Česonis, J., & Franklin, D. W. (2022). Contextual cues are not unique for motor learning: Task-dependant switching of feedback controllers. *PLOS Computational Biology*, *18*(6), e1010192. https://doi.org/10.1371/journal.pcbi.1010192

Christou, A. I., Miall, R. C., McNab, F., & Galea, J. M. (2016). Individual differences in explicit and implicit visuomotor learning and working memory capacity. *Scientific Reports*, *6*(1), 36633. https://doi.org/10.1038/srep36633

Classen, J., Liepert, J., Wise, S. P., Hallett, M., & Cohen, L. G. (1998). Rapid Plasticity of Human Cortical Movement Representation Induced by Practice. *Journal of Neurophysiology*, *79*(2), 1117–1123. https://doi.org/10.1152/jn.1998.79.2.1117

Codol, O., Kashefi, M., Forgaard, C. J., Galea, J. M., Pruszynski, J. A., & Gribble, P. L. (2023). Sensorimotor feedback loops are selectively sensitive to reward. *ELife*, *12*. https://doi.org/10.7554/eLife.81325

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 933–942. https://doi.org/10.1098/rstb.2007.2098

Cunningham, H. A. (1989). Aiming error under transformed spatial mappings suggests a structure for visual-motor maps. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 493–506. https://doi.org/10.1037/0096-1523.15.3.493

Day, K. A., Roemmich, R. T., Taylor, J. A., & Bastian, A. J. (2016). Visuomotor learning generalizes around the intended movement. *Eneuro*, *3*(2), ENEURO.0005-16.2016. https://doi.org/10.1523/ENEURO.0005-16.2016

de Leva, P. (1996). Adjustments to Zatsiorsky-Seluyanov's segment inertia parameters. *Journal of Biomechanics*, *29*(9), 1223–1230. https://doi.org/10.1016/0021-9290(95)00178-6

Diedrichsen, J., White, O., Newman, D., & Lally, N. (2010). Use-Dependent and Error-Based Learning of Motor Behaviors. *Journal of Neuroscience*, *30*(15), 5159–5166. https://doi.org/10.1523/JNEUROSCI.5406-09.2010

Doemges, F., & Rack, P. M. (1992). Task-dependent changes in the response of human wrist joints to mechanical disturbance. *The Journal of Physiology*, *447*(1), 575–585. https://doi.org/10.1113/jphysiol.1992.sp019019

Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Networks*, *12*(7–8), 961–974. https://doi.org/10.1016/S0893-6080(99)00046-5

Doya, K. (2000). Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current Opinion in Neurobiology*, *10*(6), 732–739. https://doi.org/10.1016/S0959-4388(00)00153-7

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience*, *9*(4), 292–303. https://doi.org/10.1038/nrn2258

Falconer, K., & Winter, D. A. (1985). Quantitative assessment of co-contraction at the ankle joint in walking. *Electromyography and Clinical Neurophysiology*, *25*(2–3), 135–149.

Finley, J. M., Bastian, A. J., & Gottschall, J. S. (2013). Learning to be economical: the energy cost of walking tracks motor adaptation. *The Journal of Physiology*, *591*(4), 1081–1095. https://doi.org/10.1113/jphysiol.2012.245506

Flash, T., & Hogan, N. (1985). The coordination of arm movements: an experimentally confirmed mathematical model. *The Journal of Neuroscience*, *5*(7), 1688–1703. https://doi.org/10.1523/JNEUROSCI.05-07-01688.1985

Forano, M., & Franklin, D. W. (2020). Timescales of motor memory formation in dual-adaptation. *PLOS Computational Biology*, *16*(10), e1008373. https://doi.org/10.1371/journal.pcbi.1008373

Forano, M., Schween, R., Taylor, J. A., Hegele, M., & Franklin, D. W. (2021). Direct and indirect cues can enable dual adaptation, but through different learning processes. *Journal of Neurophysiology*, *126*(5), 1490–1506. https://doi.org/10.1152/jn.00166.2021

Franklin, D. W., Burdet, E., Peng Tee, K., Osu, R., Chew, C.-M., Milner, T. E., & Kawato, M. (2008). CNS learns stable, accurate, and efficient movements using a simple algorithm. *Journal of Neuroscience*, *28*(44), 11165–11173. https://doi.org/10.1523/JNEUROSCI.3099-08.2008

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

Franklin, S., Wolpert, D. M., & Franklin, D. W. (2012). Visuomotor feedback gains upregulate during the learning of novel dynamics. *Journal of Neurophysiology*, *108*(2), 467–478. https://doi.org/10.1152/jn.01123.2011

Fusi, S., Asaad, W. F., Miller, E. K., & Wang, X.-J. (2007). A Neural Circuit Model of Flexible Sensorimotor Mapping: Learning and Forgetting on Multiple Timescales. *Neuron*, *54*(2), 319–333. https://doi.org/10.1016/j.neuron.2007.03.017

Gandolfo, F., Mussa-Ivaldi, F. A., & Bizzi, E. (1996). Motor learning by field approximation. *Proceedings of the National Academy of Sciences*, *93*(9), 3843–3846. https://doi.org/10.1073/pnas.93.9.3843

Greenberg, I., Chow, Y., Ghavamzadeh, M., & Mannor, S. (2022). Efficient risk-averse reinforcement learning. *Advances in Neural Information Processing Systems*, *35*, 32639–32652.

Hadjiosif, A. M., Krakauer, J. W., & Haith, A. M. (2021). Did We Get Sensorimotor Adaptation Wrong? Implicit Adaptation as Direct Policy Updating Rather than Forward-Model-Based Learning. *The Journal of Neuroscience*, *41*(12), 2747–2761. https://doi.org/10.1523/JNEUROSCI.2125-20.2021

Haith, A. M., & Krakauer, J. W. (2013a). *Model-Based and Model-Free Mechanisms of Human Motor Learning* (pp. 1–21). https://doi.org/10.1007/978-1-4614-5465-6_1

Haith, A. M., & Krakauer, J. W. (2013b). Theoretical models of motor control and motor learning. In *Routledge handbook of motor control and motor learning* (pp. 16–37).

Harris, C. M., & Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature*, *394*(6695), 780–784. https://doi.org/10.1038/29528

Heald, J. B., Franklin, D. W., & Wolpert, D. M. (2018). Increasing muscle co-contraction speeds up internal model acquisition during dynamic motor learning. *Scientific Reports*, *8*(1), 16355. https://doi.org/10.1038/s41598-018-34737-5

Heald, J. B., Lengyel, M., & Wolpert, D. M. (2021). Contextual inference underlies the learning of sensorimotor repertoires. *Nature*, *600*(7889), 489–493. https://doi.org/10.1038/s41586-021-04129-3

Hegele, M., & Heuer, H. (2010). Implicit and explicit components of dual adaptation to visuomotor rotations. *Consciousness and Cognition*, *19*(4), 906–917. https://doi.org/10.1016/j.concog.2010.05.005

Heuer, H., & Hegele, M. (2008). Adaptation to visuomotor rotations in younger and older adults. *Psychology and Aging*, *23*(1), 190–202. https://doi.org/10.1037/0882-7974.23.1.190

Hirashima, M., & Nozaki, D. (2012). Distinct motor plans form and retrieve distinct motor memories for physically identical movements. *Current Biology*, *22*(5), 432–436. https://doi.org/10.1016/j.cub.2012.01.042

Hogan, N., & Sternad, D. (2009). Sensitivity of Smoothness Measures to Movement Duration, Amplitude, and Arrests. *Journal of Motor Behavior*, *41*(6), 529–534. https://doi.org/10.3200/35-09-004-RC

Holland, P., Codol, O., & Galea, J. M. (2018). Contribution of explicit processes to reinforcement-based motor learning. *Journal of Neurophysiology*, *119*(6), 2241–2255. https://doi.org/10.1152/jn.00901.2017

Howard, I. S., Ingram, J. N., Franklin, D. W., & Wolpert, D. M. (2012). Gone in 0.6 seconds: The encoding of motor memories depends on recent sensorimotor states. *Journal of Neuroscience*, *32*(37), 12756–12768. https://doi.org/10.1523/JNEUROSCI.5909-11.2012

Howard, I. S., Wolpert, D. M., & Franklin, D. W. (2013). The effect of contextual cues on the encoding of motor memories. *Journal of Neurophysiology*, *109*(10), 2632–2644. https://doi.org/10.1152/jn.00773.2012

Howard, I. S., Wolpert, D. M., & Franklin, D. W. (2015). The value of the follow-through derives from motor learning depending on future actions. *Current Biology*, *25*(3), 397–401. https://doi.org/10.1016/j.cub.2014.12.037

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

Humphries, M. (2012). Dopaminergic control of the exploration-exploitation trade-off via the basal ganglia. *Frontiers in Neuroscience*, *6*. https://doi.org/10.3389/fnins.2012.00009

Ikegami, T., Ganesh, G., Gibo, T. L., Yoshioka, T., Osu, R., & Kawato, M. (2021). Hierarchical motor adaptations negotiate failures during force field learning. *PLOS Computational Biology*, *17*(4), e1008481. https://doi.org/10.1371/journal.pcbi.1008481

Imamizu, H., Uno, Y., & Kawato, M. (1995). Internal representations of the motor apparatus: Implications from generalization in visuomotor learning. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(5), 1174–1198. https://doi.org/10.1037/0096-1523.21.5.1174

Ito, M., & Doya, K. (2011). Multiple representations and algorithms for reinforcement learning in the cortico-basal ganglia circuit. *Current Opinion in Neurobiology*, *21*(3), 368–373. https://doi.org/10.1016/j.conb.2011.04.001

Izawa, J., Rane, T., Donchin, O., & Shadmehr, R. (2008). Motor Adaptation as a Process of Reoptimization. *Journal of Neuroscience*, *28*(11), 2883–2891. https://doi.org/10.1523/JNEUROSCI.5359-07.2008

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

Jordan, M. I. (1996). *Chapter 2 Computational aspects of motor control and motor learning* (pp. 71–120). https://doi.org/10.1016/S1874-5822(06)80005-8

Karniel, A., & Mussa-Ivaldi, F. A. (2002). Does the motor control system use multiple models and context switching to cope with a variable environment? *Experimental Brain Research*, *143*(4), 520–524. https://doi.org/10.1007/s00221-002-1054-4

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, *9*(6), 718–727. https://doi.org/10.1016/S0959-4388(99)00028-8

Kawato, M., & Cortese, A. (2021). From internal models toward metacognitive AI. *Biological Cybernetics*, *115*(5), 415–430. https://doi.org/10.1007/s00422-021-00904-7

Kawato, M., Furukawa, K., & Suzuki, R. (1987). A hierarchical neural-network model for control and learning of voluntary movement. *Biological Cybernetics*, *57*(3), 169–185. https://doi.org/10.1007/BF00364149

Kawato, M., & Samejima, K. (2007). Efficient reinforcement learning: computational theories, neuroscience and robotics. *Current Opinion in Neurobiology*, *17*(2), 205–212. https://doi.org/10.1016/j.conb.2007.03.004

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin*, *136*(5), 849–874. https://doi.org/10.1037/a0019842

Kim, H. E., Avraham, G., & Ivry, R. B. (2021). The Psychology of Reaching: Action Selection, Movement Implementation, and Sensorimotor Learning. *Annual Review of Psychology*, *72*(1), 61–95. https://doi.org/10.1146/annurev-psych-010419-051053

Kim, H. E., Parvin, D. E., & Ivry, R. B. (2019). The influence of task outcome on implicit motor learning. *ELife*, *8*. https://doi.org/10.7554/eLife.39882

Kluzik, J., Diedrichsen, J., Shadmehr, R., & Bastian, A. J. (2008). Reach Adaptation: What Determines Whether We Learn an Internal Model of the Tool or Adapt the Model of Our Arm? *Journal of Neurophysiology*, *100*(3), 1455–1464. https://doi.org/10.1152/jn.90334.2008

Komar, J., Potdevin, F., Chollet, D., & Seifert, L. (2019). Between exploitation and exploration of motor behaviours: unpacking the constraints-led approach to foster nonlinear learning in physical education. *Physical Education and Sport Pedagogy*, *24*(2), 133–145. https://doi.org/10.1080/17408989.2018.1557133

Krakauer, J. W. (2009). Motor learning and consolidation: The case of visuomotor rotation. *Advances in Experimental Medicine and Biology*, *629*, 405–421. https://doi.org/10.1007/978-0-387-77064-2_21

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*, *93*(3), 480–490. https://doi.org/10.1016/j.neuron.2016.12.041

Krakauer, J. W., Hadjiosif, A. M., Xu, J., Wong, A. L., & Haith, A. M. (2019). Motor Learning. In *Comprehensive Physiology* (pp. 613–663). Wiley. https://doi.org/10.1002/cphy.c170043

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

Krutky, M. A., Ravichandran, V. J., Trumbower, R. D., & Perreault, E. J. (2010). Interactions Between Limb and Environmental Mechanics Influence Stretch Reflex Sensitivity in the Human Arm. *Journal of Neurophysiology*, *103*(1), 429–440. https://doi.org/10.1152/jn.00679.2009

Kunavar, T., Cheng, X., Franklin, D. W., Burdet, E., & Babič, J. (2023). Explicit learning based on reward prediction error facilitates agile motor adaptations. *PLOS ONE*, *18*(12), e0295274. https://doi.org/10.1371/journal.pone.0295274

Lackner, J. R., & Dizio, P. (1994). Rapid adaptation to Coriolis force perturbations of arm trajectory. *Journal of Neurophysiology*, *72*(1), 299–313. https://doi.org/10.1152/jn.1994.72.1.299

Langsdorf, L., Maresch, J., Hegele, M., McDougle, S. D., & Schween, R. (2021). Prolonged response time helps eliminate residual errors in visuomotor adaptation. *Psychonomic Bulletin & Review*, *28*(3), 834–844. https://doi.org/10.3758/s13423-020-01865-x

Lee, D., Seo, H., & Jung, M. W. (2012). Neural Basis of Reinforcement Learning and Decision Making. *Annual Review of Neuroscience*, *35*(1), 287–308. https://doi.org/10.1146/annurev-neuro-062111-150512

Lee, J.-Y., & Schweighofer, N. (2009). Dual Adaptation Supports a Parallel Architecture of Motor Memory. *The Journal of Neuroscience*, *29*(33), 10396–10404. https://doi.org/10.1523/JNEUROSCI.1294-09.2009

Lee, K., Oh, Y., Izawa, J., & Schweighofer, N. (2018). Sensory prediction errors, not performance errors, update memories in visuomotor adaptation. *Scientific Reports*, *8*(1), 16483. https://doi.org/10.1038/s41598-018-34598-y

Leow, L., Marinovic, W., de Rugy, A., & Carroll, T. J. (2018). Task errors contribute to implicit aftereffects in sensorimotor adaptation. *European Journal of Neuroscience*, *48*(11), 3397–3409. https://doi.org/10.1111/ejn.14213

Leow, L.-A., Marinovic, W., de Rugy, A., & Carroll, T. J. (2020). Task Errors Drive Memories That Improve Sensorimotor Adaptation. *The Journal of Neuroscience*, *40*(15), 3075–3088. https://doi.org/10.1523/JNEUROSCI.1506-19.2020

Maia, T. V, & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience*, *14*(2), 154–162. https://doi.org/10.1038/nn.2723

Maurus, P., Mahdi, G., & Cluff, T. (2024). Increased muscle coactivation is linked with fast feedback control when reaching in unpredictable visual environments. *IScience*, *27*(11), 111174. https://doi.org/10.1016/j.isci.2024.111174

Mawase, F., Uehara, S., Bastian, A. J., & Celnik, P. (2017). Motor Learning Enhances Use-Dependent Plasticity. *The Journal of Neuroscience*, *37*(10), 2673–2685. https://doi.org/10.1523/JNEUROSCI.3303-16.2017

Mazzoni, P., & Krakauer, J. W. (2006a). An Implicit Plan Overrides an Explicit Strategy during Visuomotor Adaptation. *The Journal of Neuroscience*, *26*(14), 3642–3645. https://doi.org/10.1523/JNEUROSCI.5317-05.2006

Mazzoni, P., & Krakauer, J. W. (2006b). An Implicit Plan Overrides an Explicit Strategy during Visuomotor Adaptation. *The Journal of Neuroscience*, *26*(14), 3642–3645. https://doi.org/10.1523/JNEUROSCI.5317-05.2006

McDougle, S. D., Bond, K. M., & Taylor, J. A. (2015). Explicit and Implicit Processes Constitute the Fast and Slow Processes of Sensorimotor Learning. *Journal of Neuroscience*, *35*(26), 9568–9579. https://doi.org/10.1523/JNEUROSCI.5061-14.2015

McDougle, S. D., Ivry, R. B., & Taylor, J. A. (2016). Taking Aim at the Cognitive Side of Learning in Sensorimotor Adaptation Tasks. *Trends in Cognitive Sciences*, *20*(7), 535–544. https://doi.org/10.1016/j.tics.2016.05.002

Merel, J., Botvinick, M., & Wayne, G. (2019). Hierarchical motor control in mammals and machines. *Nature Communications*, *10*(1), 5489. https://doi.org/10.1038/s41467-019-13239-6

Miall, R. C., & Wolpert, D. M. (1996). Forward Models for Physiological Motor Control. *Neural Networks*, *9*(8), 1265–1279. https://doi.org/10.1016/S0893-6080(96)00035-4

Milner, T. E., & Franklin, D. W. (2005). Impedance control and internal model use during the initial stage of adaptation to novel dynamics in humans. *The Journal of Physiology*, *567*(2), 651–664. https://doi.org/10.1113/jphysiol.2005.090449

Miyamoto, Y. R., Wang, S., & Smith, M. A. (2020). Implicit adaptation compensates for erratic explicit strategy in human motor learning. *Nature Neuroscience*, *23*(3), 443–455. https://doi.org/10.1038/s41593-020-0600-3

Morimoto, J., & Doya, K. (2001). Acquisition of stand-up behavior by a real robot using hierarchical reinforcement learning. *Robotics and Autonomous Systems*, *36*(1), 37–51. https://doi.org/10.1016/S0921-8890(01)00113-0

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

Nowak, M. A., & Sigmund, K. (2004). Evolutionary Dynamics of Biological Games. *Science*, *303*(5659), 793–799. https://doi.org/10.1126/science.1093411

Nowak, M., & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature*, *364*(6432), 56–58. https://doi.org/10.1038/364056a0

Nozaki, D., Kurtzer, I., & Scott, S. H. (2006). Limited transfer of learning between unimanual and bimanual skills within the same limb. *Nature Neuroscience*, *9*(11), 1364–1366. https://doi.org/10.1038/nn1785

O'Connor, S. M., & Kuo, A. D. (2009). Direction-Dependent Control of Balance During Walking and Standing. *Journal of Neurophysiology*, *102*(3), 1411–1419. https://doi.org/10.1152/jn.00131.2009

Oh, Y., & Schweighofer, N. (2019). Minimizing Precision-Weighted Sensory Prediction Errors via Memory Formation and Switching in Motor Adaptation. *The Journal of Neuroscience*, *39*(46), 9237–9250. https://doi.org/10.1523/JNEUROSCI.3250-18.2019

Orban de Xivry, J.-J., & Lefèvre, P. (2016). A switching cost for motor planning. *Journal of Neurophysiology*, *116*(6), 2857–2868. https://doi.org/10.1152/jn.00319.2016

Osu, R., Franklin, D. W., Kato, H., Gomi, H., Domen, K., Yoshioka, T., & Kawato, M. (2002). Short- and Long-Term Changes in Joint Co-Contraction Associated With Motor Learning as Revealed From Surface EMG. *Journal of Neurophysiology*, *88*(2), 991–1004. https://doi.org/10.1152/jn.2002.88.2.991

Palidis, D. J., Cashaback, J. G. A., & Gribble, P. L. (2019). Neural signatures of reward and sensory error feedback processing in motor learning. *Journal of Neurophysiology*, *121*(4), 1561–1574. https://doi.org/10.1152/jn.00792.2018

Peacock, K. A. (2011). The three faces of ecological fitness. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *42*(1), 99–105. https://doi.org/10.1016/j.shpsc.2010.11.011

Pekny, S. E., Izawa, J., & Shadmehr, R. (2015). Reward-Dependent Modulation of Movement Variability. *The Journal of Neuroscience*, *35*(9), 4015–4024. https://doi.org/10.1523/JNEUROSCI.3244-14.2015

Posch, M. (1999). Win–Stay, Lose–Shift Strategies for Repeated Games—Memory Length, Aspiration Levels and Noise. *Journal of Theoretical Biology*, *198*(2), 183–195. https://doi.org/10.1006/jtbi.1999.0909

Pruszynski, J. A., Kurtzer, I., Lillicrap, T. P., & Scott, S. H. (2009). Temporal Evolution of "Automatic Gain-Scaling." *Journal of Neurophysiology*, *102*(2), 992–1003. https://doi.org/10.1152/jn.00085.2009

Reimann, H., Fettrow, T. D., Thompson, E. D., Agada, P., McFadyen, B. J., & Jeka, J. J. (2017). Complementary mechanisms for upright balance during walking. *PLOS ONE*, *12*(2), e0172215. https://doi.org/10.1371/journal.pone.0172215

Rohrer, B., Fasoli, S., Krebs, H. I., Hughes, R., Volpe, B., Frontera, W. R., Stein, J., & Hogan, N. (2002). Movement Smoothness Changes during Stroke Recovery. *The Journal of Neuroscience*, *22*(18), 8297–8304. https://doi.org/10.1523/JNEUROSCI.22-18-08297.2002

Sainburg, R. L., Ghez, C., & Kalakanis, D. (1999). Intersegmental Dynamics Are Controlled by Sequential Anticipatory, Error Correction, and Postural Mechanisms. *Journal of Neurophysiology*, *81*(3), 1045–1056. https://doi.org/10.1152/jn.1999.81.3.1045

Samejima, K., Ueda, Y., Doya, K., & Kimura, M. (2005). Representation of Action-Specific Reward Values in the Striatum. *Science*, *310*(5752), 1337–1340. https://doi.org/10.1126/science.1115270

Scheidt, R. A., Reinkensmeyer, D. J., Conditt, M. A., Rymer, W. Z., & Mussa-Ivaldi, F. A. (2000). Persistence of Motor Adaptation During Constrained, Multi-Joint, Arm Movements. *Journal of Neurophysiology*, *84*(2), 853–862. https://doi.org/10.1152/jn.2000.84.2.853

Schultz, W. (2017). Reward prediction error. *Current Biology*, *27*(10), R369–R371. https://doi.org/10.1016/j.cub.2017.02.064

Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Schween, R., Langsdorf, L., Taylor, J. A., & Hegele, M. (2019). How different effectors and action effects modulate the formation of separate motor memories. *Scientific Reports*, *9*(1), 17040. https://doi.org/10.1038/s41598-019-53543-1

Schween, R., McDougle, S. D., Hegele, M., & Taylor, J. A. (2020). Assessing explicit strategies in force field adaptation. *Journal of Neurophysiology*, *123*(4), 1552–1565. https://doi.org/10.1152/jn.00427.2019

Schween, R., Taylor, J. A., & Hegele, M. (2018). Plan-based generalization shapes local implicit adaptation to opposing visuomotor transformations. *Journal of Neurophysiology*, *120*(6), 2775–2787. https://doi.org/10.1152/jn.00451.2018

Scott, S. H. (2004). Optimal feedback control and the neural basis of volitional motor control. *Nature Reviews Neuroscience*, *5*(7), 532–545. https://doi.org/10.1038/nrn1427

Sedaghat-Nejad, E., Herzfeld, D. J., & Shadmehr, R. (2019). Reward Prediction Error Modulates Saccade Vigor. *The Journal of Neuroscience*, *39*(25), 5010–5017. https://doi.org/10.1523/JNEUROSCI.0432-19.2019

Seymour, B. (2019). Pain: A Precision Signal for Reinforcement Learning and Control. *Neuron*, *101*(6), 1029–1041. https://doi.org/10.1016/j.neuron.2019.01.055

Shadmehr, R. (2004). Generalization as a behavioral window to the neural mechanisms of learning internal models. *Human Movement Science*, *23*(5), 543–568. https://doi.org/10.1016/j.humov.2004.04.003

'Shadmehr, R., 'Brashers-Krug, T., & 'Mussa-Ivaldi, F. A. (1994). Interference in learning internal models of inverse dynamics in humans. In G. Tesauro, D. Touretzky, & T. Leen (Eds.), *Advances in Neural Information Processing Systems*. MIT Press.

Shadmehr, R., & Krakauer, J. W. (2008). A computational neuroanatomy for motor control. *Experimental Brain Research*, *185*(3), 359–381. https://doi.org/10.1007/s00221-008-1280-5

Shadmehr, R., & Mussa-Ivaldi, F. (1994). Adaptive representation of dynamics during learning of a motor task. *The Journal of Neuroscience*, *14*(5), 3208–3224. https://doi.org/10.1523/JNEUROSCI.14-05-03208.1994

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

Sheahan, H. R., Franklin, D. W., & Wolpert, D. M. (2016). Motor planning, not execution, separates motor memories. *Neuron*, *92*(4), 773–779. https://doi.org/10.1016/j.neuron.2016.10.017

Shen, T., & Dayan, P. (2024). *Risking your Tail: Modeling Individual Differences in Risk-sensitive Exploration using Bayes Adaptive Markov Decision Processes*. https://doi.org/10.1101/2024.01.07.574574

Smith, M. A., Ghazizadeh, A., & Shadmehr, R. (2006). Interacting adaptive processes with different timescales underlie short-term motor learning. *PLoS Biology*, *4*(6), e179. https://doi.org/10.1371/journal.pbio.0040179

Spampinato, D., & Celnik, P. (2021). Multiple motor learning processes in humans: Defining their neurophysiological bases. *The Neuroscientist*, *27*(3), 246–267. https://doi.org/10.1177/1073858420939552

Summerside, E. M., Kram, R., & Ahmed, A. A. (2018). Contributions of metabolic and temporal costs to human gait selection. *Journal of The Royal Society Interface*, *15*(143), 20180197. https://doi.org/10.1098/rsif.2018.0197

Sutton, R. S., & Barto, A. G. (1998). Reinforcement Learning: An Introduction. *IEEE Transactions on Neural Networks*, *9*(5), 1054–1054. https://doi.org/10.1109/TNN.1998.712192

Taylor, J. A., & Ivry, R. B. (2011). Flexible Cognitive Strategies during Motor Learning. *PLoS Computational Biology*, *7*(3), e1001096. https://doi.org/10.1371/journal.pcbi.1001096

Taylor, J. A., & Ivry, R. B. (2014). Cerebellar and prefrontal cortex contributions to adaptation, strategies, and reinforcement learning. *Progress in Brain Research*, *210*, 217–253. https://doi.org/10.1016/B978-0-444-63356-9.00009-1

Taylor, J. A., Krakauer, J. W., & Ivry, R. B. (2014). Explicit and implicit contributions to learning in a sensorimotor adaptation task. *Journal of Neuroscience*, *34*(8), 3023–3032. https://doi.org/10.1523/JNEUROSCI.3619-13.2014

Tee, K. P., Franklin, D. W., Kawato, M., Milner, T. E., & Burdet, E. (2010). Concurrent adaptation of force and impedance in the redundant muscle system. *Biological Cybernetics*, *102*(1), 31–44. https://doi.org/10.1007/s00422-009-0348-z

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

Thoroughman, K. A., & Shadmehr, R. (2000). Learning of action through adaptive combination of motor primitives. *Nature*, *407*(6805), 742–747. https://doi.org/10.1038/35037588

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

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2009). The Expected Utility of Movement. In *Neuroeconomics* (pp. 95–111). Elsevier. https://doi.org/10.1016/B978-0-12-374176-9.00008-7

Truong, C., Ruffino, C., Crognier, A., Paizis, C., Crognier, L., & Papaxanthis, C. (2023). Error-based and reinforcement learning in basketball free throw shooting. *Scientific Reports*, *13*(1), 499. https://doi.org/10.1038/s41598-022-26568-2

Tsay, J. S., Haith, A. M., Ivry, R. B., & Kim, H. E. (2022). Interactions between sensory prediction error and task error during implicit motor learning. *PLOS Computational Biology*, *18*(3), e1010005. https://doi.org/10.1371/journal.pcbi.1010005

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

van Dooren, R., de Kleijn, R., Hommel, B., & Sjoerds, Z. (2021). The exploration-exploitation trade-off in a foraging task is affected by mood-related arousal and valence. *Cognitive, Affective, & Behavioral Neuroscience*, *21*(3), 549–560. https://doi.org/10.3758/s13415-021-00917-6

van Mastrigt, N. M., Smeets, J. B. J., & van der Kooij, K. (2020). Quantifying exploration in reward-based motor learning. *PLOS ONE*, *15*(4), e0226789. https://doi.org/10.1371/journal.pone.0226789

van Mastrigt, N. M., Tsay, J. S., Wang, T., Avraham, G., Abram, S. J., van der Kooij, K., Smeets, J. B. J., & Ivry, R. B. (2023). Implicit reward-based motor learning. *Experimental Brain Research*, *241*(9), 2287–2298. https://doi.org/10.1007/s00221-023-06683-w

Vaswani, P. A., Shmuelof, L., Haith, A. M., Delnicki, R. J., Huang, V. S., Mazzoni, P., Shadmehr, R., & Krakauer, J. W. (2015). Persistent Residual Errors in Motor Adaptation Tasks: Reversion to Baseline and Exploratory Escape. *The Journal of Neuroscience*, *35*(17), 6969–6977. https://doi.org/10.1523/JNEUROSCI.2656-14.2015

Wada, Y., Kawabata, Y., Kotosaka, S., Yamamoto, K., Kitazawa, S., & Kawato, M. (2003). Acquisition and contextual switching of multiple internal models for different viscous force fields. *Neuroscience Research*, *46*(3), 319–331. https://doi.org/10.1016/S0168-0102(03)00094-4

Weightman, M., Brittain, J.-S., Miall, R. C., & Jenkinson, N. (2022). Residual errors in visuomotor adaptation persist despite extended motor preparation periods. *Journal of Neurophysiology*, *127*(2), 519–528. https://doi.org/10.1152/jn.00301.2021

Wolpert, D. M., Diedrichsen, J., & Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience*, *12*(12), 739–751. https://doi.org/10.1038/nrn3112

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology*, *11*(18), R729–R732. https://doi.org/10.1016/S0960-9822(01)00432-8

Wolpert, D. M., & Flanagan, J. R. (2010). Motor learning. *Current Biology*, *20*(11), R467–R472. https://doi.org/10.1016/j.cub.2010.04.035

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An Internal Model for Sensorimotor Integration. *Science*, *269*(5232), 1880–1882. https://doi.org/10.1126/science.7569931

Wolpert, D. M., & Landy, M. S. (2012). Motor control is decision-making. *Current Opinion in Neurobiology*, *22*(6), 996–1003. https://doi.org/10.1016/j.conb.2012.05.003

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, *2*(9), 338–347. https://doi.org/10.1016/S1364-6613(98)01221-2

Wu, Y., Morita, M., & Izawa, J. (2022). Reward prediction errors, not sensory prediction errors, play a major role in model selection in human reinforcement learning. *Neural Networks*, *154*, 109–121. https://doi.org/10.1016/j.neunet.2022.07.002

Yang, Y., & Lisberger, S. G. (2010). Learning on Multiple Timescales in Smooth Pursuit Eye Movements. *Journal of Neurophysiology*, *104*(5), 2850–2862. https://doi.org/10.1152/jn.00761.2010

Bibliography

Publications Related to the Thesis

Journal Articles

Babič, J., Kunavar, T., Oztop, E., & Kawato, M. (2025). Success-efficient/failure-safe strategy for hierarchical reinforcement motor learning. *PLOS Computational Biology*, 21(5), e1013089. https://doi.org/10.1371/journal.pcbi.1013089 (shared first authorship)

Kunavar, T., Cheng, X., Franklin, D. W., Burdet, E., & Babič, J. (2023). Explicit learning based on reward prediction error facilitates agile motor adaptations. *PLOS ONE*, *18*(12), e0295274. https://doi.org/10.1371/journal.pone.0295274

Kunavar, T., Franklin, D. W., & Babič, J. (2026). Error signal type shapes strategy use in force-field motor adaptation. Manuscript submitted for publication.

Kunavar, T., & Babič, J. (2026). Adaptive control model of human motor adaptation integrating internal models and strategy learning. Manuscript submitted for publication.

Other Publications

Journal Articles

Arditi, E., Kunavar, T., Amirshirzad, N., Ugur, E., Babič, J., & Oztop, E. (2025). Inferring effort-safety trade off in perturbed squat-to-stand task by reward parameter estimation. *Engineering Applications of Artificial Intelligence, 142*(12). https://doi.org/10.1016/j.engappai.2024.109778

Kunavar, T., Kroflič, N., Jamšek, M., & Babič, J. (2025). Influence of video gaming experience on motor adaptation during visuomotor and force field perturbations. *Heliyon,* 11(10), e43109, 1-9. https://doi.org/10.1016/j.heliyon.2025.e43109

Kunavar, T., Jamšek, M., Ávila Mireles, E. J., Rueckert, E., Peternel, L., & Babič, J. (2024). The effects of different motor teaching strategies on learning a complex motor task. *Sensors, 24(*4), 1-17. https://doi.org/10.3390/s24041231

Jamšek, M., Kunavar, T., Blohm, G., Nozaki, D., Papaxanthis, C., White, O., & Babič, J. (2021). Effects of simulated microgravity and hypergravity conditions on arm movements in normogravity. *Frontiers in Neural Circuits*, *15*, 750176-1-750176-15. https://doi.org/10.3389/fncir.2021.750176 (shared first authorship)

Kunavar, T., Jamšek, M., Barbiero, M., Blohm, G., Nozaki, D., Papaxanthis, C., White, O., & Babič, J. (2021). Effects of local gravity compensation on motor control during altered environmental gravity. *Frontiers in Neural Circuits, 15*, 750267-1-750267-13. https://doi.org/10.3389/fncir.2021.750267 (shared first authorship)

Jamšek, M., Kunavar, T., Bobek, U., Rueckert, E., & Babič, J. (2021). Predictive exoskeleton control for arm-motion augmentation based on probabilistic movement primitives combined with a flow controller. *IEEE Robotics and Automation Letters*, *6*(3), 4417-4424. https://doi.org/10.1109/LRA.2021.3068892

Conference Papers

Arditi, E., Kunavar, T., Ugur, E., Babič, J., & Oztop, E. (2021). Inferring cost functions using reward parameter search and policy gradient reinforcement learning. In *IECON 2021 - 47th Annual Conference of the IEEE Industrial Electronics Society*, 1-6. IEEE. https://doi.org/10.1109/IECON48115.2021.9589967

Rottmann, N., Kunavar, T., Babič, J., Peters, J., & Rueckert, E. (2020). Learning hierarchical acquisition functions for Bayesian optimization. In *2020 IEEE/RSJ IEEE International Conference on Intelligent Robots and Systems (IROS)*, 5490-5496. IEEE. https://doi.org/5490-5496. 10.1109/IROS45743.2020.9341335

Biography

Tjaša Kunavar, MSc, is a PhD student at the Jožef Stefan International Postgraduate School and a research assistant at the Laboratory for Neuromechanics and Biorobotics at the Jožef Stefan Institute. She obtained her BSc degree in Biopsychology from the University of Primorska in 2015 and her MSc degree in Computational Neuroscience and Cognitive Robotics from the University of Birmingham in 2017.

Her research focuses on human motor learning and motor control, with particular emphasis on the interaction between sensorimotor adaptation and reward-based learning. Her work combines experimental studies of human movement with computational modelling approaches inspired by reinforcement learning to better understand how humans adapt their movements in dynamic environments.

During her doctoral studies she has been involved in several interdisciplinary research projects in collaboration with international partners and has contributed to multiple scientific publications and conference presentations in the fields of neuroscience, robotics, and motor control.