Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Imperial College London

Department of Information Technology and
Electrical Engineering at ETH Zurich
Departement of Bioengineering at Imperial College
London

Autumn 2023

Computational Model of Motor Learning in a Real-World Task

Master Thesis

Dario Bolli
dbolli@ethz.ch

November 23, 2023

Supervisor: Prof. Aldo Faisal, a.faisal@imperial.ac.uk
Dr. Shlomi Haar, s.haar@imperial.ac.uk
Professor: Valerio Mante, valerio@ini.uzh.ch

Abstract

In this project, we investigate the underlying processes driving sensorimotor learning in real-world settings.

A behavioural analysis of both a conventional and a more realistic motor adaptation task is carried out under two different feedback conditions.
A thorough review of the literature is conducted, with a focus on learning based on sensory prediction error and reward feedback. Then, two computational models are proposed to explain the underlying mechanisms contributing to the experimentally observed behaviour.
The models are subsequently fitted to each of the participants in the experiment, demonstrating their ability to capture several aspects of sensorimotor adaptation including exponential learning, suboptimality, between and within participants' motor variability, as well as reward-dependent modulation of the variability in the case of reward feedback.
Based on anatomical evidence and experimentally verified assumptions, these models provide a foundation for advancing research in sensorimotor adaptation and can be readily extended to other motor adaptation tasks.

In conclusion, potential avenues for future research are suggested to enhance our understanding of motor learning mechanisms and their implications for neurodegenerative diseases.

The code can be found upon request at [1].

Declaration of Originality

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Authored by (in block letters):
For papers written by groups the names of all authors are required.

Name(s):	**First name(s):**

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	**Signature(s)**

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.

Dario Bolli, London, November 23, 2023

Contents

Chapter 1

Introduction

1.1. Motor Learning

Motor learning refers to the intricate process of acquiring and refining motor skills through changes occurring in the nervous system, by relying on sensory feedback. This essential characteristic of mammals and more generally all living beings occurs across various temporal scales and levels of complexity. From mastering fundamental skills like walking and talking over the course of several years to continuously adapting to changes in weight, height, and strength in their daily life, humans engage in a lifetime journey of motor learning. This adaptive process contributes to the fluidity and precision of movements, allowing individuals to acquire new skills applicable to a wide range of contexts, from everyday tasks such as reaching for an apple to more complex activities including learning a new sport or the rehabilitation of motor function after a stroke.

Learning a motor skill in real-world scenarios is generally prolonged, intricate, and challenging to quantify. This complexity arises from the integration of motor skills across different tasks. For instance, mastering basketball involves incorporating previously acquired motor skills like running, throwing, and catching a ball.
Moreover, consistent repetition of the same movement is likely to yield different outcomes each time. The variations observed between those consecutive trials are called motor variability.
This variability stems from the diverse ways in which a redundant musculoskeletal system can coordinate joints, muscles, and limbs to accomplish the same task[2]. Additionally, the presence of noise at different stages in the motor processes further contributes to the varying nature of the outcomes[3].

Consequently, disentangling the intricate processes at play and identifying the underlying neural mechanisms resulting from a specific movement or behaviour presents a

challenging task.

Depending on the task at hand, motor learning involves several processes including sensory integration (e.g. visual, auditory, somatosensory, and proprioceptive feedback), explicit decision making, coordination, biomechanical control and adaptation [4][5].
To better understand these mechanisms, researchers have proposed various experimental settings aimed at studying them in isolation.
In this project, our emphasis is on sensorimotor adaptation, a fundamental aspect of motor learning involving the refinement of pre-existing motor skills to align with changes in environmental circumstances or in the human body.

A well-established method to study this phenomenon consists in introducing a perturbation in a procedural task. This has been investigated in highly controlled and simplified laboratory experiments allowing a well-defined, restricted and readily observable framework to study motor adaptation. Typically, the task consists of hand-reaching movements towards a target[6], and the perturbation is introduced in the form of a force-field [7] or a visuomotor rotation [8][9].

Visuomotor rotation is a specific case of visuomotor adaptation task where the learner observes her movements indirectly (i.e. through a cursor on a digital screen) and does not see his arm and hand. The cursor movement is shifted by an angular rotation and a discrepancy between the visual feedback and the expected sensory feedback leads the learner to adjust the movements.

In a typical visuomotor rotation task, the perturbation is introduced as a rotation on the trajectory of the displayed cursor, and the participant must learn to correct it to still reach the target.

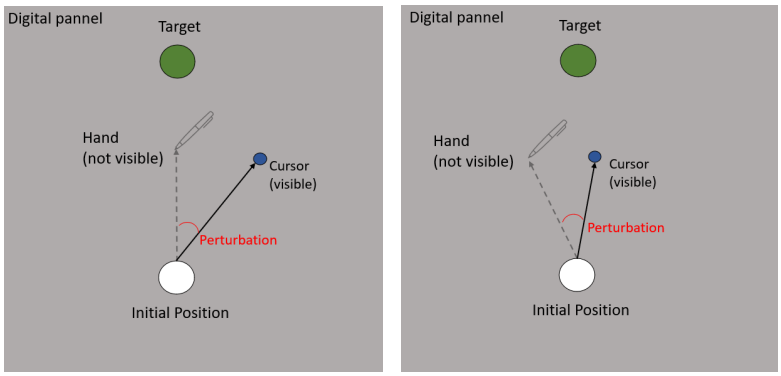


Figure 1.1.: Typical Baseline (left) and asymptotic (right) trajectories of a participant when the perturbation is introduced

1. Introduction

An altered behaviour following the introduction of a perturbation can arise from an explicit or an implicit process [10].
In the explicit process, a conscious, high-level strategy, informed by the observation of performance errors or external explicit verbal or visual cues, results in a deliberate adjustment of motor commands, akin to continuous decision-making.
Conversely, in the implicit mechanism, motor commands are unconsciously altered when a disparity is noted between the expected sensory feedback and the actual outcome feedback (e.g. hand position).

It has been demonstrated that adaptation to a visuomotor rotation is learned implicitly and occurs independently of explicit learning, even in the presence of a conflicting cognitive strategy. Additionally, any observed discrepancy between expectations and reality triggers a motor adaptation process aimed at minimising this prediction error[8].

Two sources of information in the visual feedback are driving an implicit adaptation without engaging an explicit strategy [4][6][11]. First, there is the difference between the anticipated visual consequence of a motor command and its actual result (irrespective of the task's objective), called the sensory prediction error [12][13].
Additionally, akin to the Reinforcement Learning field, the task's success or failure information is sufficient to drive behaviour[14] without the need for an explicit strategy. It has been demonstrated that participants were able to adapt to a visuomotor rotation implicitly relying solely on reward feedback while showing no awareness of the perturbation [15]. The discrepancy between the predicted reward of a motor command and the binary outcome observed is referred to as a reward prediction error.

It has been observed [11] in a traditional visuomotor reaching task that participants were relying predominantly on the sensory error feedback. However, as the quality of this visual feedback declined, a shift towards a predominant reliance on reward feedback was observed, thus highlighting the essential role of the two forms of feedback information.

Furthermore, it has been suggested that learning through these two different forms of feedback depends on separate physiological mechanisms [16].
A challenge in the motor adaptation field is to understand the underlying mechanisms processing those different information signals.
The neural basis underlying motor learning has been primarily identified in the Cortex, specifically the primary motor cortex (M1) and primary somatosensory cortex (S1), as well as in the cerebellum and Basal Ganglia.
The observation of motor adaptation deficits following impairments in the Cerebellum [17] and in the Basal Ganglia [18] emphasises their crucial role in sensorimotor adaptation. Drawing on their distinct anatomical structures, it has been proposed that the cerebellum relies on sensory prediction errors, while the basal ganglia rely on reward prediction error to drive the adaptation process[19].

So far, these behavioural principles of sensorimotor adaptation have been studied in highly restrictive laboratory experiments only, limiting their generalisation to a real-world scenario.

Although current experiments offer valuable insights into the neurobehavioural dynamics and play a vital role in our understanding of motor adaptation, their application is limited to a narrow range of behaviours, which may not fully reflect the complexity of real-world scenarios.
Hence, to gain a more comprehensive understanding of how the brain assimilates feedback information and adjusts its behaviour accordingly, it is imperative to observe this phenomenon in a real-world context.

Chapter 2

Previous work

2.1. Pool task

To investigate this paradigm in a real-world scenario, S. Haar and A. A. Faisal proposed a motor learning task that reflects real-world challenges in the form of learning to play pool[20].
The game of billiards has well-defined objectives and is readily observable. On top of that, it requires complex skills and involves several sub-tasks such as precision, alignment, ballistic movements, or high-level sequential planning of shots and ball positions. Consequently, billiards serve as an interesting paradigm for investigating motor learning within a real-world, complex task.
During the study, participants received instructions to use a cue stick to execute a pool shot, aiming to pocket a target ball in a specific pocket on the billiard table. The participants had the freedom to move their entire body, engaging in self-paced movements without any imposed restrictions.
During the experiment, high-speed camera tracking was employed to monitor the positions of the balls and the cue stick and participants wore a motion tracking "suit" containing wireless Inertial Measurement Unit sensors to capture their body movements. Mobile brain imaging was utilised to monitor EEG activity. This experimental setup enabled participants to execute genuine motor commands and perceive natural somatosensory feedback within a realistic environment.

Neural signatures of motor learning are significant in the neural beta oscillations (13–30Hz)[21], and specifically in the post-movement beta rebound (PMBR)[22]. The PMBR is a temporary change in beta oscillation amplitude observed in the sensorimotor cortex following the end of a voluntary movement. This effect is believed to reflect the involvement of the sensorimotor network in motor planning, execution and feedback processing, and has been observed in a variety of motor tasks. In the context of motor adaptation tasks,

opposing patterns in the beta activity have been observed for reward-based and sensory error-based motor learning[23][24].

In the pool task, the analysis of the PMBR dynamics revealed two distinct groups of participants exhibiting opposing dynamics[25]. These findings indicated the presence of different predominant learning mechanisms for the same motor adaptation task[26].
The authors demonstrate that engaging in real-world motor learning, characterised by fewer constraints on the tasks, provides individuals greater flexibility in using diverse learning processes. Consequently, this freedom may potentially lead to different learning strategies among individuals.

Furthermore, based on the previously observed association between different types of implicit adaptation and PMBR, they proposed that when performing the same task under identical conditions, one group of participants relies more on the reward prediction error, whereas the other predominantly utilises the sensory prediction error to drive the adaptation process.

To allow further investigation of this hypothesis, S. Haar and A. A. Faisal devised a method to manipulate the feedback provided to human subjects while preserving a sense of realism by the means of Embodied Virtual Reality (EVR) [27]. The participant wore embodied virtual reality glasses and could observe the environment (i.e. cue ball, target ball, billiard table and cue stick) indirectly through the glasses. This setup allowed us to manipulate the visual feedback. Similar to the classical visuomotor rotation task, an angular rotation is introduced on the trajectory of the cue ball. This perturbation creates a discrepancy between the learner's visual expectations and what is actually perceived, prompting him to make corrections to successfully pocket the target ball.

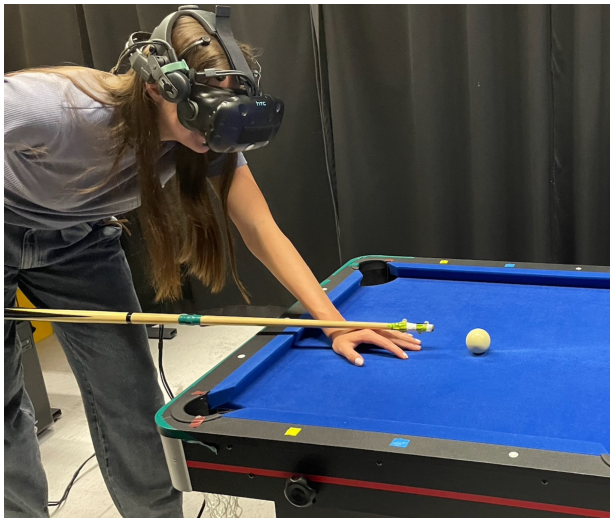


Figure 2.1.: Embodied Virtual Reality in the pool task [28]

Expanding on the EVR pool task setup, F. Nardi, M. Ziman, S. Haar, and A. A. Faisal [28] proposed two experiments to isolate and investigate the error-based and reward-based learning mechanisms in a real-world setting with the help of forty participants. By using a Virtual Reality headset, the researchers were able to manipulate the visual feedback provided to the subjects when introducing a visuomotor perturbation, effectively enforcing the use of either one of the specific adaptation mechanisms.
The perturbation consisted of a 5° counterclockwise rotation of the cueball trajectory in the virtual environment, thus forcing the participant to modify his shot to pocket the ball (see fig [2.2]).

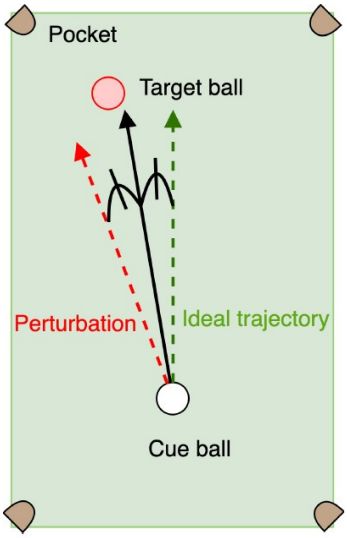


Figure 2.2.: perturbation of the cueball trajectory in the virtual environment

In the error-based feedback group, the trajectories of the balls were shown until the collision between the cue ball and the target ball only (i.e. The subsequent trajectory of the target ball toward the pocket was hidden). This constrains the subjects to improve their performance based on the error of the cue ball's trajectory towards the target ball, without receiving any information on the performance or success of the trial.
Following the introduction of a perturbation, participants observe a disparity between the anticipated trajectory of the ball and the observed rotated trajectory. Consequently, they adjust their motor commands to align the observed trajectory with the anticipated trajectory.

In the reward-based feedback group, the actual trajectories of the balls were not displayed. The participants were shown an artificial successful trajectory (always identical) in case of a successful outcome. Based on statistical analysis, the reward zone was defined as a -0.31, $+0.25$ margin around the optimal angle for pocketing.
Given that adaptation to an abrupt perturbation might cause the participant to give up exploration in the absence of rewards, a dynamic reward zone gradually guides the participant toward the optimal angle (i.e. the angle the participants need to aim at to pocket the target ball while accounting for the 5° perturbations) was introduced. This adaptive reward zone was calculated as the zone between the median of the past ten rewarded trials, and the lower margin of the optimal angle accounting for perturbations. This type of feedback does not allow any estimation of the magnitude or direction of the error but provides information on the success or failure of the trial.
In the reward feedback case, participants learn the optimal motor commands that maximise task performance. When the perturbation is introduced, the motor commands associated with a successful outcome change, and participants engage in the process of

relearning the optimal commands.

Each participant performed both the error-based and reward-based tasks. To account for a potential bias due to the order in which the tasks are performed and the perturbation directions, the forty subjects were divided into four groups performing the tasks in different orders with different combinations of feedback and perturbation direction.
Each participant completed ten blocks of twenty-five shot trials with a set break in between to mitigate muscle fatigue. The initial three blocks served as a baseline phase, allowing participants to familiarise themselves with the task and learn to pocket the ball with full visual feedback. Following this, a visual rotation was introduced in the subsequent six blocks with task-specific feedback, and finally, the last block was used as a washout block to observe the after-effects of the perturbation.
This experiment showed that in a realistic task, successful learning can be achieved through either error-based or reward-based learning mechanisms individually.
The adaptation behaviour of the participant can be characterised by the change in the cue ball angle.
After introducing a rotation in the visual feedback, the angle relative to the target (i.e. the optimal cue ball angle for pocketing the target ball) shows two distinct behaviours depending on the provided feedback.

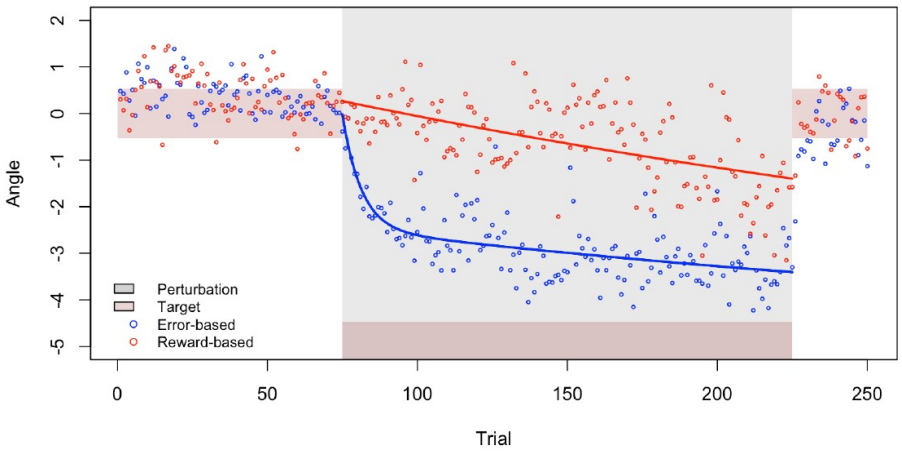


Figure 2.3.: angle adaptation

2.2. Hand Reaching task

To compare sensorimotor adaptation in a controlled laboratory setting, a more conventional experiment was conducted by H. Li and S. Haar using a hand-reaching task[29]. A total of twenty right-handed participants (12 men and 8 women) were recruited. The task involved making repetitive forward-reaching movements using an electrical pen on

a digital panel, aiming for a displayed target without visual feedback from their hand.

In the error-feedback condition, subjects were provided with visual feedback in the form of a cursor on the screen, representing the position of the digital pen during the reaching movement, up to the point where the trajectory reached half the screen's length. In the reward-based condition, the trajectory was hidden, and participants received rewards on the screen upon successful reaching.

The reward zone was defined as a 5° margin on each side of the target. During the perturbation phase, a 15° perturbation was introduced, causing a misalignment between the participant's hand movement and the cursor's movement on the screen. Feedback to the participant was based on the cursor's position, either as an error or a reward.
Throughout the perturbation phase, the reward zone was progressively adapted to drive the participant towards the optimal angle to account for the 15° perturbations. This dynamic reward zone was calculated as the zone between the median of the past 10 trials, and the lower bound of the optimal angle.
The participants' objective was to learn how to counteract the rotation by reaching for the angle opposite to the target. Each participant performed both the error-based and reward-based tasks. To counterbalance the sensitivity to the order of feedback tasks and perturbation directions, the twenty subjects were divided into four groups performing the tasks in different orders with different combinations of feedback and perturbation direction. Brain activity was recorded using a high-density EEG system with 256 channels, and hand movements were recorded by a digital panel and stored on a computer.
Once more, each participant completed twelve blocks of twenty-five reaching trials with a set break in between. The initial two blocks served as a baseline phase, the subsequent eight blocks as the adaptation phase, and the final block as a washout block. As the task is considerably simpler, all the trials were performed with task-specific feedback.

2.3. Semester Project

Our previous work proposed to apply state-of-the-art reinforcement learning to model the learning of the pool task's policy in the reward feedback case.
Specifically, we proposed two different Off-policy Actor-Critic architectures trained on the dataset collected in the pool experiment.
Off-policy learning provides a robust framework for acquiring a policy through reinforcement learning without direct interaction with the environment. Given sufficient data, it effectively uncovers the optimal policy.

However, several challenges emerge when attempting to model the adaptation behaviour using the optimal RL agent.
In comparison to participants, the RL agent demonstrates lower sensitivity to changes

in the reward. The perturbation is treated as a new task, and due to the substantial size of the neural networks in the actor and the critic, thousands of examples are needed to alter the agent's behaviour.
Indeed, deep reinforcement learning relies on having an extensive dataset to extract meaningful patterns and develop an optimal policy.
Conversely, the adaptation behaviour exhibited by participants in the experiments shows an instantaneous alteration of the behaviour, and the adaptation effect is observed on a much shorter timescale (100 trials).

Additionally, a key characteristic of motor learning is that two consecutive attempts at repeating the same movement inevitably result in somewhat different outcomes [2]. Noise is not an undesirable side effect, rather, it holds a fundamental role in motor learning, particularly when relying solely on reward feedback (see [3.3]). This consideration prompts us to explore an approach akin to what has been proposed in the motor adaptation literature [6][4][30][11].

Most importantly, while it is crucial to introduce computational models that accurately capture observations, the objective of such models is to provide a potential explanation of the underlying mechanisms. With this in mind, simpler models offer clearer explanations that are typically more robust for generalisation and require minimal assumptions.

Consequently, in this project, following the literature in the sensorimotor adaptation field [3.3], we explore simpler alternatives based on the actor-critic framework and the reward prediction error.
Those models make minimal assumptions and are grounded in neurophysiological and anatomical evidence.

2.4. Motivation

In this project, our objective is to explore computational approaches grounded in biological evidence that capture and explain the data observed in a realistic visuomotor adaptation task.
This work aims at gaining insights into the underlying computational mechanisms driving implicit sensorimotor adaptation and its application to real-world scenarios.

Our work is motivated by an intriguing real-life application, which is to ease the detection and address the rehabilitation of neurodegenerative diseases affecting the cerebellum and the basal ganglia, such as cerebellar stroke, cerebellar ataxia, Huntington's, and Parkinson's disease.

More specifically, individuals affected by Parkinson's disease experience neural impairments in the substantia nigra, a region within the basal ganglia [31], where reward-based learning is thought to occur [6][32][33].
Therefore, a better understanding of the processes at play in the basal ganglia during sensorimotor adaptation could help address the detection and potential rehabilitation of Parkinson's disease.

Chapter 3

Theory

Planning a movement towards a target involves the integration of information on the limb position and target location in the external task space along with its transformation to an internal representation, composed of time-varying sets of muscle activations and joint torques[9].

It has been suggested that at higher levels, the nervous system represents and plans reaching movements in the extrinsic space [2][8][34][35].
Therefore, sensorimotor adaptation is typically observed in the extrinsic space by looking at the changes in angles between successive movements directed toward a target [9][12][36][37].
Throughout the remainder of the report, we will use the terms "reach angle" (i.e. reaching task), "angle" and "cue ball angle" (i.e. pool task) interchangeably to refer to the angle characterising the adaptation behaviour.

The observed variations in motor outcomes stem from diverse sources of noise within the motor system [3], along with an alteration of the motor command following perceived sensory feedback.
In the following section, we will review the existing theory associated with each form of feedback, examining the adaptation process through the integration of noisy sensory feedback.

3.1. Learning from sensory prediction error

To successfully act and adapt within a dynamic environment, we continuously monitor sensory feedback associated with our movements and accordingly alter our motor commands such that our actions yield the desired outcomes.
Our ability to anticipate the outcome of an action suggests the existence of an internal model [38] predicting the consequences of a motor command.
An internal representation of our movements arises from the integration of the proprioceptive and visual feedback information[39][40].
The difference between the anticipated consequence and the internal representation of the perceived sensory feedback is the sensory prediction error and has been shown to drive sensorimotor adaptation[12][38][41][42][43].
In contrast with performance error observed on the task outcome, sensory prediction error denotes the difference between the predicted motor consequences and the actual observed sensory consequences.
In a standard visuomotor reaching task, the performance error is determined by the difference between the endpoint of the movement and the target, whereas sensory prediction error manifests during the reaching movement and involves a comparison between the anticipated hand position and the actual hand position.
Typically, to observe the behaviour of the participant following a sensory prediction error feedback, the hand position is displayed only up to half of the movement's trajectory, so that it does not allow the estimation of the performance error [10][12].

This internal forward model cannot plausibly be a simple association between each action and a resulting consequence given the number of different combinations of muscles and limbs in the musculoskeletal system.
Therefore, it has been proposed that the brain approximates an unknown dynamical system representing the processes intervening between motor commands and desired outcomes (i.e. the musculoskeletal system and associated noise, the physics of the environment, and the sensory receptors). This functional approximation relies on a forward model recursively updated by the sensory feedback[41][42][44].

The computation of the sensory prediction error is believed to take place in the cerebellum[45] through the use of the forward internal model[46].
It was observed that patients suffering from cerebellar impairments (i.e. Cerebellar ataxia) were unable to correctly predict the expected motor consequence during a visuomotor reaching task[17].

More specifically, a supervised learning framework has been introduced[42]. In this model, the cerebellum's forward model translates an efferent copy of descending commands [47] into an estimate of the dynamic state of the system allowing the prediction of the anticipated sensory feedback. Then, the computation of the sensory prediction error is used to update the forward model.

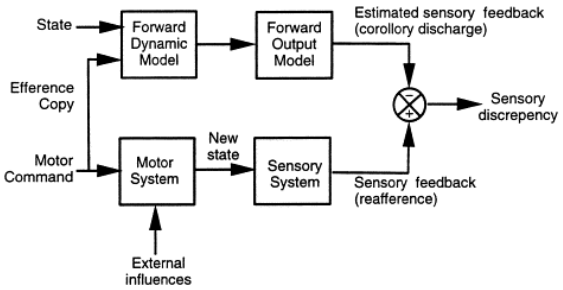


Figure 3.1.: internal forward model[42]

The existence of the forward internal model is biologically plausible [19][41] and is supported by behavioural and neural activity [48][49][50].
Specifically, neurophysiological evidence suggests that the activity of Purkinje cells in the cerebellum reflects the resulting movement outcome[51], and precedes the actual true state of the limbs[52]. These findings provide support for the existence of an internal forward model that calculates the anticipated consequences of motor commands in the extrinsic space.

Muscle fatigue, bone growth, muscle mass increase, or diseases can affect musculoskeletal dynamics. Furthermore, changes in the environment such as the weight, velocity or force exerted on an object we are interacting with, can have an impact on the outcome of an action. Hence, continuous recalibration of the forward model is essential.
When observing a discrepancy in the sensory prediction error, the forward model realigns the internal estimate of the system's dynamic and new motor commands can be computed accordingly.

From the learner's perspective, errors may arise due to random noise present at various stages.
Within the motor system, randomness originates from the accumulation of noise at different levels. This includes synaptic noise resulting from the inherent variability in chemical processes during synaptic transmission, electrical noise in neurons arising from fluctuations in voltage-gated ion channels, and cellular noise due to stochastic biochemical processes (i.e. protein production, synaptic vesicle fusion, diffusion and binding of signalling molecules to receptors)[3].

Additionally, variability can emerge from uncertainties in the feedback signal due to noise in sensory signals, sensory receptors, or environmental factors. Although not precise, the aggregation of these random sources can be approximated using a Gaussian distribution[3].
Alternatively, errors can stem from a shift in the relationship between a motor command

and the movement outcome, as both the body and the environment are subject to potential changes.
Hence, the appropriate correction following the observation is not evident, as the system should respond differently if the error is due to a systematic change or to random noise.

Consider a scenario where a pool player executes consecutive shots to pocket a target ball in an identical setting. In one instance, the target ball misses the pocket and strikes the band 5cm too far to the right. To rectify this, the player might adjust by aiming 5cm more to the left on the subsequent shot.
Alternatively, one could argue that the aiming direction was correct, but random noise during the execution of the movement caused the error. In this case, it might not be necessary for the player to modify the planned trajectory for the next shot.
Another perspective could be to adopt an optimal strategy by aiming at a point in-between to mitigate the impact of error.
This simple example highlights the challenge of determining the source of the error and shows the importance of noise in this process.

An approach to this problem is to rely on past knowledge of the dynamical system and incorporate new observations weighted by their uncertainty to update the estimate of the system's dynamics. This has been addressed originally by Bayesian inference in probability theory [3.1.2].
It has been observed that participants behaviour in several perception and adaptation experiments was consistent with a Bayesian observer [53][54][55][56] [57]. The Bayesian approach allows to weight the update of the forward model by the uncertainty in the sensory feedback and the state estimate.
Furthermore, studies have demonstrated that the distribution of endpoints in a procedural reaching task aligns well with a Gaussian distribution [58] [59].

Therefore, using an analogy in the field of control engineering, where we want to estimate the state variables describing a dynamic system given noisy measurements, the Kalman filter [3.1.3] has been proposed to update the forward model optimally given the diverse noise sources[60].

The Kalman filter is used to calculate the optimal update of the state of the dynamical system given the prior state and the sensory feedback affected by noise[60][61].

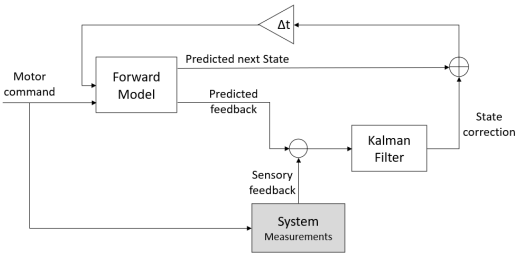


Figure 3.2.: Kalman filter to update the forward model [61]

This approach has been applied successfully in the context of motor adaptation to qualitatively capture the movement adaptation in conventional reaching task [37][54].

Moreover, it has been shown [62] that the degrading reliability of the sensory feedback was leading to a slower adaptation rate, whereas increased uncertainty in the state estimation was leading to a faster adaptation rate, further supporting the Bayesian view.

The biological plausibility of the Kalman filter in the cerebellum has been suggested [63]. More specifically, the authors observed that the Purkinje cell implements a filtering function such that its output converges to the desired response by minimising the mean squared error.

Historically, the selection of a motor command has been thought to originate from an inverse model provided with an optimal trajectory[49].

More recently, stochastic optimal control feedback has been proposed to address motor control[64]. This method uses Linear Quadratic Gaussian control to determine the optimal control input of the dynamical system minimising a specific motor cost function, based on the estimated underlying state [3.1.4].
As introduced earlierKalman-sensory-estimation, the forward model, updated via the Kalman filter, estimates the system's state.
As a result of the separation principle of estimation and control, the tasks of estimating the model dynamics and generating the optimal motor command can be independently addressed[65].

This framework adjusts the motor output based on variations in the estimated state created by external perturbations and system noise.
In contrast to classical approaches [49], planning of specific trajectories is not required. Correcting for the goal-relevant variability only, while leaving the task-irrelevant variations in state variables uncorrected allows greater flexibility in the movement[66].
This control strategy, called the minimal intervention principle has been proven to be optimal when facing uncertainty in a dynamical system[67].
Moreover, this is in line with the various ways in which a redundant musculoskeletal sys-

tem can combine joints, muscles and limbs to achieve the same task[2], and the so-called uncontrolled manifold theory [34], that enforce the importance of controlling the degrees of freedom relevant to the task while permitting more variability in the unrelated degrees of freedom.
Finally, it is also consistent with several aspects of neural processing in the primary motor cortex M1 [68] and has been supported by behavioural evidence [69].

A widely used framework for describing dynamical systems is provided by State Space Models.

3.1.1. State Space Model

State-space models (SSMs) allow the expression of the temporal dynamics of a system through the evolution of a partially observable variable called state.
This framework allows the characterisation of systems that may be continuous or discrete-time, deterministic or stochastic, time-variant or invariant, and linear or nonlinear.
The dynamics of the system are determined by the unobservable latent state x, which evolves over time according to a predefined set of equations, and the input u to the system. The input-output relationship is characterised by another set of equations and depends on this latent state.
The combination of those two equations fully characterises the system. This general formulation allows the characterisation of many signals and systems.

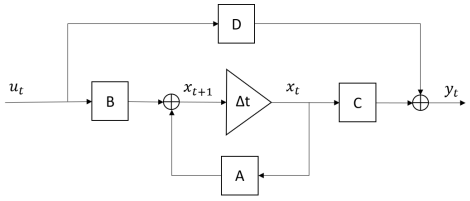
It can be expressed in its most general form for continuous-time non-linear systems as follows:

$$\dot{x} = f(t, x(t), u(t))$$
$$y = h(t, x(t), u(t)$$

To characterise the adaptation as a dynamical system we use the discrete-time, linear formulation which can be expressed as:

$$x[n+1] = Ax[n] + Bu[n]$$
$$y[n] = Cx[n] + Du[n]$$

Where A corresponds to the state transition matrix, B stands for the input matrix, C corresponds to the observation matrix, and D stands for the feedforward matrix.



The matrix formulation allows to express the entangled relationship between different state variables (such as position, velocity and acceleration).
State space models provide a unified flexible framework for modelling and analysing complex systems. They enable us to characterise dynamical systems with multiple interacting components, and sources of uncertainty. These models have demonstrated their effectiveness in understanding and predicting the behaviour of diverse real-world phenomena and have been widely used in fields such as engineering, statistics, economics, and computational neuroscience.

In neuroscience experiments, measurements of neural or behavioural data frequently exhibit noise and intricate temporal dynamics. State space models have been introduced as statistical instruments in computational neuroscience to capture the inherent dynamic characteristics present in the neural and behavioural responses of experimental participants.

3.1.2. Bayesian Inference

Based on Baye's theorem [70], Bayesian inference describes how the probability of a hypothesis evolves with the accumulation of additional information.
The probability of a hypothesis H is based on prior knowledge about the hypothesis, the likelihood of the observed data given that hypothesis $P(E|H)$, and the observed data E (also called evidence):

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \tag{3.1}$$

3.1.3. Kalman Filter

The Kalman Filter is an algorithm designed to estimate the underlying state of a dynamical system in the presence of noise.
It comes originally from the field of optimal control and is widely used in engineering applications such as target tracking, signal processing, navigation, and control, but also in time series analysis, statistics and econometrics.

Building on State-Space models, Kalman introduced an optimal state estimator for Linear dynamical systems [71]. This estimator is optimal in the sense that it minimises the mean squared error between the model prediction and the observations at each timestep.

Kalman filter characterises the state of a dynamical system by a posterior probability distribution over the state variables. This posterior distribution is calculated based on the prior estimation of the distribution along with the integration of new observations. The new observations allow us to compute an error between the output of the system based on prior beliefs and the observed measurements. This error is weighted by the uncertainty present at each stage in the system and used to update the state estimate.

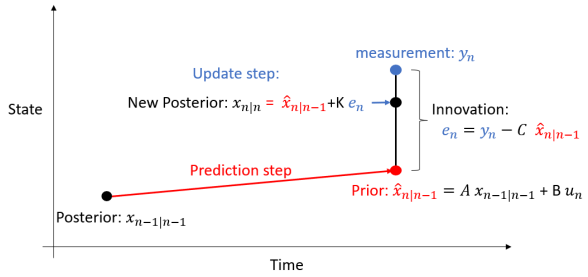


Figure 3.3.: Kalman step, adapted from [72]

Generally, the Kalman filter is constructed based on a linear state space model that assumes normally distributed noise:

$$\begin{aligned} x_{n+1} &= A_n x_n + B_n u_n + w_n, \quad w_n \sim \mathcal{N}(0, W_n) \\ \tilde{y}_n &= C_n xn + v_n, \quad v_n \sim \mathcal{N}(0, V_n) \end{aligned} \tag{3.2}$$

We call w_n the process noise, and v_n the observation noise.
The matrices W_n and V_n are diagonal matrices of shape consistent with the dimension of the state x, with the process noise $\sigma^2_{w,n}$ and the measurement uncertainty $\sigma^2_{v,n}$ respectively on the diagonal.
To characterise the state at timestep n, the Kalman filter uses a posterior distribution, with mean $x_{n|n}$, and covariance matrix $P_{n|n}$, giving a measure of the uncertainty of the state estimate.
This distribution is computed iteratively, relying on the prior estimation of the distribution $x_{n|n-1}$, $P_{n|n-1}$ and integrating the new observations made at timestep n.
The computation of the estimated prior distribution is known as the prediction step. The posterior distribution of the previous timestep $n-1$ is passed through the state space model to compute the prior estimate of the current timestep n:

$$\begin{aligned} \hat{x}_{n|n-1} &= A_n x_{n-1|n-1} + B_n u_n \\ \hat{P}_{n|n-1} &= A_n P_{n-1|n-1} A_n^T + W_n \end{aligned} \tag{3.3}$$

The discrepancy between the measurements and the predicted output of the system based on the prior estimate is called the innovation term.

$$\tilde{e}_n = \tilde{y}_n - C_n \hat{x}_{n|n-1}$$

To update the estimated state distribution, this error is weighted by the process and observation uncertainties and integrated into the prior estimate of the state.
This is known as the update step.
The innovation covariance S_n estimates the system uncertainty by projecting the covariance matrix $\hat{P}_{n|n-1}$ in the measurement space.
The Kalman gain K_n balances the state estimate based on how much trust the filter places in the measurements versus in the process model. Looking at [3.3], K_n defines where the posterior state estimate lies on the line between the prior and the observation. This gain computes the ratio of the uncertainty in the prior against the uncertainty in the measurements and is dynamically adjusted at each time step as the uncertainties evolve.

This can be seen in the one-dimensional state case where the Kalman gain is given by:

$$K = \frac{\sigma_w^2}{\sigma_w^2 + \sigma_v^2}$$

Using the Kalman gain K_n, we can update the posterior state and covariance estimate:

$$\begin{aligned} S_n &= C_n \hat{P}_{n|n-1} C_n^T + V_n \\ K_n &= \hat{P}_{n|n-1} C_n^T S_n^{-1} \\ x_{n|n} &= \hat{x}_{n|n-1} + K_n \tilde{e}_n \\ P_{n|n} &= (I - K_n C_n) \hat{P}_{n|n-1} \end{aligned} \tag{3.4}$$

Refer to [73] for proof of the derivation.

By combining noisy measurements over time with prior information, the Kalman filter maximises the accuracy and precision of the estimate (i.e. minimises the variance and bias).
In our task, the internal forward model approximates a system composed of the body and the environment. It is expressed as a discrete-time SSM, receiving motor commands (i.e. intended angle) in input and returning sensory feedback (i.e. observed angle). The state of the system (i.e. estimation of perturbation and observed angle) is updated using a Kalman filter and sensory feedback from the environment.

The likelihood of observing the measurement y_n at time n, given the State Space Model's parameters and the previous observations is used in the update step of the Kalman filter [3.4] and is given by:

$$
\begin{aligned}
&p(y_n|y_{n-1},...,y_1,\theta) = p(y_n|x_n)p(x_n|y_{n-1},...,y_1,\theta) \\
&= \mathcal{N}(y_n; Cx_n, \sigma_m^2)\mathcal{N}(x_n; \hat{x}_{n|n-1}, P_{n|n-1}) \\
&= \mathcal{N}(y_n; C\hat{x}_{n|n-1}, CP_{n|n-1}C^T + \sigma_m^2) \\
&= \mathcal{N}(y_n; C\hat{x}_{n|n-1}, S_n)
\end{aligned}
$$

Where $\hat{x}_{n|n-1}$ and $P_{n|n-1}$ are respectively the prior state and covariance matrix estimates [3.3], and S_n is the innovation covariance matrix [3.4].

The probability of the entire sequence of N observations y_1^N given the model's parameters θ, is calculated using the chain rule. It is expressed as the product of the probability of each observation given the previous observations.
As our variables are distributed according to normal distributions, taking the log-likelihood allows us to simplify the exponential form of the distribution:

$$
\begin{aligned}
&log(\mathcal{L}(y_1^N|\theta) = log(\prod_1^N p(y_n|y_{n-1},...,y_1,\theta)) = log(\prod_{n=1}^N \mathcal{N}(y_n; C\hat{x}_{n|n-1}, S_n)) \\
&= -\frac{1}{2}\sum_{n=1}^N (y_n - \hat{y}_n)^T S_n^{-1}(y_n - \hat{y}_n) - \frac{1}{2}\sum_{n=1}^N log(|S_n|) - \frac{N}{2}log(2\pi) \\
&\text{with} \quad \hat{y}_n = C^T\hat{x}_{n|n-1}
\end{aligned}
\tag{3.5}
$$

3.1.4. Linear Quadratic Gaussian Control

In optimal control theory, a particularly interesting objective is to operate a dynamical system with the lowest possible cost.
This cost is typically defined as the deviation of the system's state from its desired value (i.e. 0° in our case). It can also factor in the magnitude of the control input, in our case the motor command.
The Linear Quadratic problem is the case when the system dynamics can be described by a set of linear differential equations and the cost is expressed as a quadratic function. The solution is then a feedback controller called a linear quadratic regulator (LQR).

This problem can be defined in both continuous and discrete time, but in our case, we are interested in the discrete-time and infinite horizon.
The SSM and its associated cost function J are defined as:

$$
\begin{aligned}
x_{n+1} &= A_n x_n + B_n u_n \\
\mathcal{J} &= \mathbb{E}[\sum_{i=0}^N x_i^T Q x_i + u_i^T R u_i]
\end{aligned}
$$

Where the matrices Q and R weigh the deviation of the system state and the motor command magnitude respectively.
The feedback gain that minimises the cost $\mathcal{J}$, depends on variables in the future. Therefore, the optimal cost to go H is defined backwards in time.

$$H_{n-1} = A_n^T H_n A_n - (A_n^T H_n B_n)(R + B_n^T H_n B_n)^{-1}(B_n^T H_n A_n) + Q$$

Alternatively, for the infinite-horizon problem, the discrete-time algebraic Riccati equation (DARE) gives a solution by eigenvalue decomposition.

$$A_n^T H A_n - H - (A_n^T H B_n)(R + B_n^T H B_n)^{-1}(B_n^T H A_n) + Q = 0$$

The optimal control law L to operate the system is given by:

$$u_n = -Lx_n$$
$$L = (R + B^T P B)^{-1} B^T P A$$

In the presence of random noise in the system, the optimal control law to suit the overall goals can be derived by combining the Kalman filter to produce a state estimate, and the Linear Quadratic regulator to derive the optimal control feedback. This approach is called Linear Quadratic Gaussian control, and one of its interesting properties is the separation principle[65], which states that the estimator (i.e. Kalman Filter) and the controller (i.e. LQR) can be designed independently.
The Mathematical proof of the derivation of the optimal feedback gain and the solution to the Riccati equations can be found here[65].

Todorov and colleagues introduced the idea of using an optimal feedback controller to model motor coordination as a dynamical system in any given task[64]. They postulate that the motor system approximates the best possible control scheme for a given task, which will generally take the form of a feedback control law.

3.2. Error is not enough

Despite previous work with computational models based on the Kalman filter [54] [74] [62] and optimal control feedback [69] being consistent with behavioural data, and successfully explaining how the motor system deals with uncertainty, some inconsistencies have been observed.

Interestingly, while cerebellar ataxia patients struggle to correctly anticipate sensory feedback, they nonetheless show the ability to adapt to the perturbation in a visuomotor

reaching task [17].

The cerebellum and basal ganglia have been recognised for their implication in the motor system based on the observation of motor deficits following damage to these structures. Doya [19], building on experimental evidence, proposed that these structures are specialised for different types of learning mechanisms. Specifically, the cerebellum implements error-based supervised learning, while the basal ganglia uses reinforcement learning.
This view is supported by numerous studies on error-based learning models, exploring notably their biological plausibility[41] and observed behavioural patterns[43].
Additionally, the reinforcement learning paradigm in the basal ganglia is strongly supported by evidence from the midbrain dopamine system and the function of dopaminergic neurons [75] in the substantia nigra, which play a crucial role in encoding the reward prediction error and the reinforcement of behaviour [76][77].

Parkinson's disease is characterised by the degeneration of dopaminergic neurons in the substantia nigra, a critical component of the basal ganglia circuitry [31][78][79].
This loss of dopaminergic neurons is thought to disrupt the reward system and affect motor learning.
Indeed, several studies have shown that despite exhibiting normal behaviour on many learning tasks, Parkinson's disease patients exhibit motor adaptation deficits in reward-related tasks[18][80][81], highlighting the importance of the basal ganglia and its reward circuitry in sensorimotor adaptation.
Additionally, studies have shown that suppressing reward information from sensory feedback during a visuomotor reaching task impairs the learning process[36].
Remarkably, despite cerebellar lesions leading to coordination and motor learning deficits, Therrien and colleagues [82] demonstrated that individuals with cerebellar ataxia could still adapt to a visuomotor perturbation task suggesting the use of a reinforcement mechanism in the basal ganglia. However, while ataxia patients displayed normal exploration variability and learned through reinforcement feedback, their high levels of motor noise restricted the extent of this learning.
Further experiments have investigated the adaptation capability of the subjects relying on sensory error or binary reinforcement feedback signal [11] [83]revealing that both types of feedback contribute to guiding motor adaptation.
Finally, it has been proposed that the two different feedback signals engage distinct underlying mechanisms[30]. The authors hypothesised that prediction errors guiding adaptation indirectly through the use of a forward model was likely to predominate in adaptation paradigms, whereas reward signal directly driving the motor commands adaptation was more likely to predominate in skill tasks[30].

The dependence on different feedback signals can result in the use of distinct neurophysiological mechanisms, as evidenced by the contrasting effects of binary and vector feedback on task performance in the cerebellar and Motor Cortex (M1) respectively[84].

Those findings suggest the existence of two distinct mechanisms processing separately the reward and the sensory prediction error feedback in the case of sensorimotor adaptation.

The nature of the information provided by the two different feedback signals and their corresponding utilisation strategies represent a fundamental difference between the two potential learning mechanisms.

In prediction error-based learning, the mappings between sensory targets and motor commands are refined trial by trial in response to sensory error feedback, relying on the recalibration of a forward internal model, allowing generalisation to unseen situations[85].

Conversely, binary feedback signals convey considerably less information and the learner must rely on a sequential exploration of the action space, either through random exploration or hypothesis generation to maximise the reward.
This exploration allows the learner to link motor commands to subjective values and a balance needs to be made between further exploring the action space or exploiting actions that lead to more favourable outcomes.

3.3. Learning from Reward prediction error

It is well-established that to learn from binary reinforcement feedback, the brain leverages dopaminergic neurons, which respond to reward prediction errors and integrate information from prior trial outcomes[76]. It uses an estimated value of the expected reward for a specific action and continually updates this value based on accumulated experience.

However, our comprehension of how humans learn from success or failure is largely based on scenarios involving a finite selection of distinct and unrelated actions, as seen typically in conventional decision-making tasks. In sensorimotor tasks, the range of possible actions is continuous and may exhibit correlations. For instance, the outcome of a chosen action can provide information about nearby actions in the motor space. If a particular motor command leads to high rewards, nearby actions also hold high value, and conversely, unrewarded actions suggest that high-reward regions are situated further away in the motor space.
Furthermore, the inherent complexity of the sensorimotor system contributes to the variability observed in the motor outcome. Each repetition of the same task results in a different neural, motor, and behavioural pattern. This variability arises from the different sources of noise in the nervous system [3], along with the multiple possible ways to combine the joints, muscles, and limbs employed to achieve the same objective[2].

Variability in movements is not only an undesirable consequence of a noisy nervous system. It allows for greater robustness in the neuronal networks [3], and it has been suggested that this variability in task-irrelevant dimensions is beneficial for learning[34] and is actively regulated[86]. Moreover, in the case of reinforcement feedback, it has been proposed that the learner actively regulates movement variability based on the reward information to explore the motor space[81]. From this perspective, variability can be viewed as a mode of exploration, signifying that the nervous system deliberately alters motor commands in an attempt to uncover actions associated with the highest reward.

Additionally, it has been shown that, in contrast to learning guided by sensory prediction error, reward-based learning does not exhibit as effective of a generalisation. This suggests the inability to perform sensory remapping through the update of the forward model when relying solely on the reward information[11].
As the error feedback signal carries sufficient information to determine the direction and magnitude of the correction needed for the optimal command, error-based models [37][64][69], do not require an exploration mechanism. On the other hand, within the reinforcement learning paradigm, there is a need for a robust exploration mechanism that actively regulates motor variability.

In a sensorimotor task, variability stems from various sources. Motor execution introduces noise, attributed to stochastic processes found at all levels in the brain[3][87][88]. Additionally, there is noise associated with the planning of motor commands, particularly resulting in a random walk in the task-irrelevant dimensions[89].
During the planning stage of a movement, neural activity, particularly in the premotor cortex, exhibits noise that corresponds to variations subsequently observed in motor behaviour[90]. It has been proposed that the motor corrections are made on the planning signal of the previous trial, suggesting that the learner is aware of this planning noise but ignorant of the execution noise[91].
Finally, a distinct structure within a basal ganglia-related circuitry that influences the variability has been identified and demonstrated to be essential for motor learning in juvenile birds [92] and it has been further suggested that dopamine might be involved in the modulation of motor variability [93].
Similarly, active modulation of variability within the basal ganglia by the human nervous system has also been observed, contributing to the facilitation of motor learning[86] [81]. This suggests that movement variability is not solely a byproduct of a noisy nervous system but also a consciously regulated mechanism that enhances the motor learning process.

Reward-based motor adaptation has been the subject of a few studies and several models attempting to explain the observed behavioural data have been proposed[11][82][83][81][94][95][96]. These models are building on the observation that random noise sources can be approximated as normally distributed [3], and the distributions of endpoints in a procedural reaching task align with a Gaussian distribution[58] [59].

They uniformly rely on a State Space Model formulation, recurrently updating a mean action (i.e. intended angle), and accounting for the variability stemming from various noise sources.
However, the specific algorithms and their behavioural consequences differ. The following studies examine the reach angle adaptation within a visuomotor adaptation task.

To enhance clarity and enable a more direct comparison between the following studies, I will standardise the notation between the models. I define the notation as follows: The angle y is determined by an internal estimate of the angle that would lead to success x (i.e. intended angle), along with motor execution noise ϵ^m, planning noise ϵ^p, and exploration noise ϵ^e.

The difference between sensory prediction error feedback and reward prediction feedback has been explored in a reaching task with manipulated visual feedback[11]. Their findings suggest that adaptation primarily depends on sensory prediction error when it is available and of high quality, but the subject can still adapt when the sensory error is either unavailable or the visual feedback is of low quality, by relying on success or failure. They proposed a model combining a Kalman filter to approximate the expected perturbation relying on the sensory prediction error and reinforcement learning to learn changes to the motor commands that maximise the reward.
Specifically, they define the reward-based model as follows:

$$
\begin{aligned}
&y_n = x_n + \epsilon_n^m, \\
&\epsilon_n^m \sim \mathcal{N}(0, \sigma_m^2) \\
&x_{n+1} = x_n + \alpha_r \delta_n \epsilon_n^m \\
&\delta_n = r_n + \gamma V_{n+1} - V_n \\
&\text{where } V_n = w_v \text{ is learnt through temporal difference learning} \\
&w_v^{n+1} = w_v^n + \alpha_v \delta_n
\end{aligned}
\tag{3.6}
$$

Where ϵ^m is some random noise used in the motor execution and the planning step.
This approach relies on an actor-critic architecture to update the value associated with an action and a policy. The authors showed that their model can qualitatively capture the behaviour of participants receiving reward feedback only in a reaching task.

Additionally, Therrien & al [11]introduced a reinforcement learning model aligning with their observations that cerebellar patients could still engage in learning by leveraging the reward feedback signal and that a higher level of motor noise was impairing the learning process.
They proposed that the learner is aware of a part of the overall movement variability, specifically the planning noise, and corrects solely on this part, as suggested by Van Beers[91]. They also simplified the temporal difference learning algorithm.
In their experimental setup, the expected reward was consistently around 50%, which

meant that every success resulted in a reward higher than expected, and every failure led to a reward lower than expected. As a result, they proposed to update the aimed angle in the direction of the planning noise only in the case of a successful trial.

$$\begin{aligned} y_n &= x_n + \epsilon_n^m + \epsilon_n^p, \\ \epsilon_n^m &\sim \mathcal{N}(0, \sigma_m^2), \quad \epsilon_n^p \sim \mathcal{N}(0, \sigma_p^2) \\ x_{n+1} &= x_n + r_n \epsilon_n^p \end{aligned} \tag{3.7}$$

The proposed model, and particularly the distinct effects of ϵ_n^m and ϵ_n^p, reflects the outcomes of their experiments. Their findings indicate that while augmenting motor noise appears to have a detrimental effect on the reinforcement learning process, increasing planning noise up to a certain threshold seems to be advantageous.

How can the learner effectively leverage this variability to explore the continuous action space?
An unrewarded trial most likely indicates that high-reward regions are further away in the motor space and conversely. An intuitive approach to modulating exploration variability would therefore be to increase the variability in motor commands after an unrewarded trial and decrease variability following a rewarded trial.

Examining the influence of rewards on variability in a visuomotor adaptation reaching task Shadmehr & al [81] observed a modulation of movement variability depending on past reinforcement signals. Specifically, they showed that variability following unrewarded trials was greater compared to rewarded ones. Moreover, beyond the influence of the current reward alone, it was noted that the history of past rewards appeared to modulate variability. Notably, this reward-induced modulation of variability was found to be compromised in Parkinson's patients, indicating a link between this mechanism and the reinforcement learning system of the Basal Ganglia.

To investigate the effect of the reward feedback signal on the movement variability, the authors defined trial-to-trial variability as the absolute difference between two consecutive reach angles y and proposed a modulation of this variability based on past rewards:

$$|\Delta y[t]| = \alpha_0(1 - r_n) + \alpha_1(1 - r_{t-1}) + \alpha_2(1 - r_{t-2}) + \epsilon$$

Upon fitting the data to subjects, they noted that $\alpha_0 > \alpha_1 > \alpha_2$, aligning with the widely expected pattern that more recent reward has a more significant impact on variability.

In a follow-up experiment, Therrien & al [88] explored the impact of varying levels of artificially introduced motor noise on healthy participants in a reaching task. This study confirmed their earlier results, and they introduced a revised model that incorporates an additional modulation of the variability based on the previous reward.

$$
\begin{aligned}
&y_n = x_n + \epsilon_n^m + \epsilon_n^p + (1 - r_{t-1})\epsilon_n^e, \\
&\epsilon_n^m \sim \mathcal{N}(0, \sigma_m^2), \quad \epsilon_n^p \sim \mathcal{N}(0, \sigma_p^2), \quad \epsilon_n^e \sim \mathcal{N}(0, \sigma_e^2) \\
&x_{n+1} = x_n + r_n[\epsilon_n^p + (1 - r_n)\epsilon_n^e]
\end{aligned} \tag{3.8}
$$

In their model, part of the planning variability is actively regulated by the participant according to the outcome of the past trial.

Another attempt to investigate motor learning based on binary reinforcement feedback signal has been proposed by Cashaback & al[83]. Specifically, they proposed that the learner uses gradient information of the reward landscape to make the most of the reward feedback in a continuous set of possible actions.
Their findings indicated that a steeper reinforcement gradient resulted in quicker learning when compared to a shallower gradient.
Moreover, when participants encountered a complex reinforcement landscape with opposing steep and shallow slopes, the steepest slope was preferred.
These observations suggest that individuals rely on local gradient information of the reward in the motor space to identify solutions that maximise reward. Additionally, they introduced a model that supports their experimental outcomes. The variability is characterised by random motor execution noise, and exploration noise conditioned on the previous reward.

$$
\begin{aligned}
&y_n = x_n + \epsilon_n^m + (1 - r_n)\epsilon_n^e, \\
&\epsilon_n^m \sim \mathcal{N}(0, \sigma_m^2), \quad \epsilon_n^e \sim \mathcal{N}(0, \sigma_e^2) \\
&x_{n+1} = x_n + r_n\alpha(y_n - x_n) = x_n + r_n\alpha[\epsilon_n^m + (1 - r_n)\epsilon_n^e]
\end{aligned} \tag{3.9}
$$

Their model is similar to [3.8], but the planning variability is entirely regulated by the participant according to the outcome of the past trial.

One of the significant challenges in neuroscience is the lack of data and the potential resulting noise in the observed dataset, making it complex to discern patterns and even more difficult to identify the influence of regulated variability.
Collecting a substantial amount of data from human subjects is a daunting task. Therefore, researchers in the motor learning field often turn to vertebrates, such as rats, monkeys and birds, which share important brain structures with humans such as the basal ganglia and the cerebellum.
In a comprehensive investigation of motor variability modulation in the basal ganglia, Smith & al[95] conducted an adaptation task experiment on rats relying on binary reinforcement feedback signals. This approach allows for the extraction of clearer patterns since the random noise inherent in motor behaviours is smoothed out through the analysis of millions of trials. While we need to be cautious when extrapolating these findings, they align with previous observations in human subjects[81].
The study proposed an extensive analysis of movement variability conditioned on reward. They determined the degree of exploratory variability following the outcome of a single

trial.
They confirmed that variability was influenced by the sequence of past rewards and found that this effect decreases exponentially as one considers trials further back in time, with an average time constant of approximately five trials.
Their results suggest that the brain computes a running estimate of the reward rate, approximating a weighted average of the past sequence of rewards, to modulate the variability.

$$
\begin{aligned}
&y_n = x_n + \epsilon_n^m + \epsilon_n^e, \\
&\epsilon_n^m \sim \mathcal{N}(0, \sigma_m^2), \quad \epsilon_n^e \sim \mathcal{N}(0, \sigma_e^2) \\
&x_{n+1} = x_n + \alpha_\mu \delta_n \epsilon_n^e, \\
&\delta_n = r_n - \bar{r}_n \\
&\bar{r}_{n+1} = \bar{r}_n + \alpha_r \delta_n
\end{aligned}
\tag{3.10}
$$

Upon further analysis exploring the connection between motor variability and the reward rate revealed an exponential relationship. Low reward rates had a substantially greater impact on variability compared to high reward rates.
While research in decision-making [97] has shown that the nervous system can alternate between two strategies to introduce variability in choices based on the reward rate, these results suggest that in the motor learning field, variability may be modulated by the reward rate more continuously.

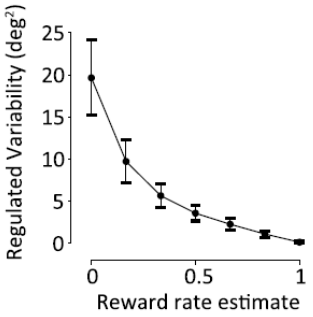


Figure 3.4.: Relationship between the regulated motor variability and the reward rate estimate [95]

Based on this observation, the authors introduced a control variability function that connects the estimated reward rate to a corresponding modulation of variability. Fitting this function to individual subjects revealed different responsiveness in the modulation of variability.

Finally, they proposed to learn an optimal control variability function relying on temporal difference learning and introduced eligibility traces to regulate the time scale over

which the variability control policy optimised future reward.

$$\begin{aligned} \sigma_{e,t}^2 &= \varsigma_n \dot{R}_n + \epsilon_n^\sigma \\ E_{n+1}^\sigma &= \lambda E_n^\sigma + \epsilon_n^\sigma \bar{R}_n \\ \varsigma_{n+1} &= \varsigma_n + \alpha_\sigma \delta_n \lambda E_n^\sigma \end{aligned} \tag{3.11}$$

This model effectively accounts for the overall observed intertrial variability pattern as well as the diverse sensitivities of the regulated variability in relation to the reward rate observed among different subjects.

Chapter 4

Data Analysis

A challenge in motor learning is to understand the mechanisms underlying sensorimotor adaptation.
To adjust his movements, the learner relies on a combination of proprioceptive and visual feedback [39]. To this end, the visuomotor adaptation task has been well-established as an effective tool for exploring the underlying mechanisms of sensorimotor adaptation.
This task typically involves introducing a rotation in the visual feedback while the participant is executing consecutive movements toward an objective. In implicit adaptation, the learner remains unaware of the perturbation and doesn't employ any cognitive strategy to counter it [10][15].

Instead, by observing a mismatch between the sensory expectations and the received sensory feedback, along with binary information on the task outcome, the learner unconsciously corrects for the perturbation.
In this project, we are exploring the use of these information signals to adjust movements in response to an introduced perturbation. Specifically, we aim at uncovering the distinct underlying mechanisms that process these two different types of feedback.
It is believed that the learner relies on the sensory prediction error, to update an internal forward model of the environment's dynamics in the cerebellum [3.1], and on the reward prediction error to reinforce successful actions in the basal ganglia [3.3].

To explore this, we investigate the data of two experiments that have been previously conducted at the brain and behaviour lab, namely a realistic pool task [2.1] and a traditional reaching task [2.2].
In those two tasks, participants were provided with manipulated visual feedback forcing them to rely solely on one feedback, either sensory error or reward.

The sensorimotor adaptation of participants to the introduced perturbation is charac-

terised at a behavioural level by the change in angles between two subsequent movements.

In the reaching task, the angle is defined as the angular difference between the vector extending from the initial position to the end position of the pen and the vector from the initial position to the optimal target location.
For the pool task, the angle is computed by considering the recorded position of the cue ball. It is determined as the angular difference between the vector extending from the initial position of the cue ball to its location approximately 110 milliseconds after being struck and the vector connecting one corner to the other on the shorter side of the pool table. To have similar quantities between the two tasks, the cue ball's angle is then referenced relative to the optimal pocketing angle.
In both tasks, the perturbation was applied in two opposite directions based on the participant's subgroup, a strategy employed to address potential biases [2]. Therefore, we invert the angles of half of the subjects to apply a comprehensive analysis of the entire dataset in each task.

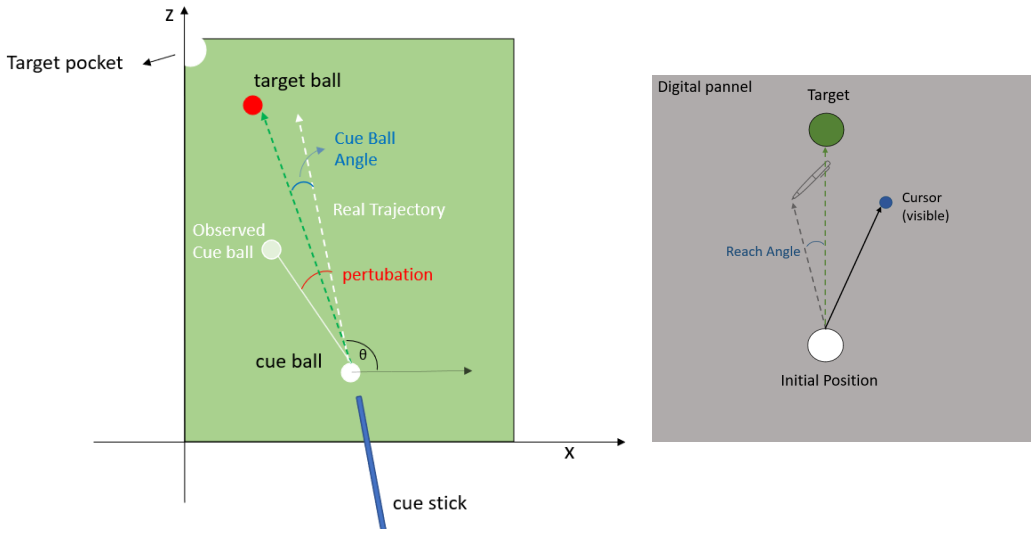


Figure 4.1.: Pool Task(left) and Reaching task (right) angle definition

The adaptation session is composed of 250 (pool task) or 300 (reaching task) consecutive angles. We visualise an adaptation session as discrete angles time-series (see [A.4] for examples).

First, we observe the average behaviour of the participants in each task to have a better understanding of the behavioural dynamics.
In the sensory error feedback condition, we anticipate observing a characteristic double-exponential learning curve during the perturbation phase[37].

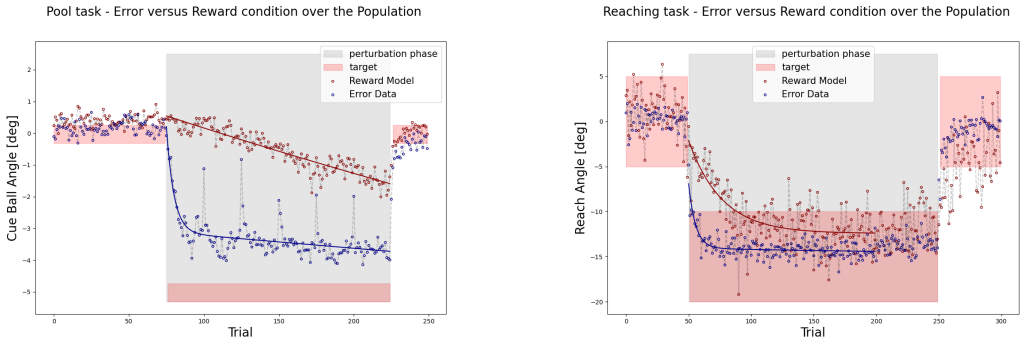


Figure 4.2.: Angles average over participants in each condition

The error condition does exhibit a double-exponential adaptation curve in both tasks and in general, participants find it more challenging to adapt to reward feedback, which is consistent with previous findings [11]. In the pool task, the adaptation is suboptimal, indicating that participants were unable to fully account for the perturbation in either of the feedback conditions. Furthermore, the learning in the reward condition seems to follow a more linear pattern.
Notably, we see the forgetting effect of the set breaks between the blocks of twenty-five trials in the pool tasks [98]. (see [A.2] for a more comprehensive block analysis of the variability and the success rate).

The different behaviours exhibited in the adaptation curves so far (as well as in the success rate and block variability [A.2]), can be explained by the two different feedbacks and task settings. However, it does not necessarily imply the existence of two distinct underlying mechanisms.

An essential difference between the two feedbacks resides in the nature of the information they convey and the subsequent strategies that can be used to update the behaviour accordingly.
As mentioned earlier [3.3], while the error feedback signal enables direct estimation of the correction required for the optimal command, in the case of reward, it is believed that exploration through regulated variability plays a fundamental role.
Therefore, we analysed the effect of a trial outcome on motor variability.

Previous studies have suggested that reward actively influences the modulation of the variability, and thus, the exploration of the motor space [3.3].
To verify this hypothesis, akin to [81] we define the intertrial movement variability as the trial-to-trial change in angle direction u_t: $|\Delta u_t| = |u_{t+1} - u_t|$. This measure provides an approximate quantification of the unsigned motor variability at each trial.

To assess the influence of reward on variability, we defined a sliding window spanning $[-10, +20]$ trials, centred around a conditioned trial (i.e., the trial where the impact of the reward condition is observed). This window facilitates the observation of intertrial variability around the conditioned trial. Subsequently, these windows are categorised into two groups based on the reward received at the conditioned trial. Averaging the windows within these two groups allows us to observe the impact of a specific reward/no reward condition on the motor variability with the window around the conditioned trial. Finally, we calculate the difference between the two averaged windows to observe the modulation effect of the conditioned trial on the intertrial variability of the movement.
In the analysis, we excluded the variability that directly depends on the conditioned trial (i.e. the intertrial variabilities that were calculated using the conditioned trial directly). Looking at a Participant example in the pool task, we notice two distinct patterns based on the reward condition.

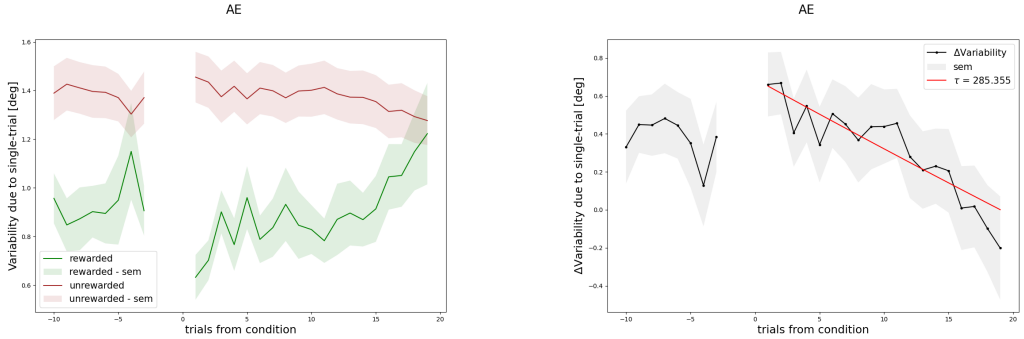


Figure 4.3.: Intertrial Variability following a conditioned trial in an example subject

While it appears that the reward condition may decrease variability, the reward-dependent modulation of the variability is not clearly distinguishable.
Additionally, the resulting effect on the window appears noisy, especially in the rewarded condition, and the number of windows is imbalanced (i.e. 191 unrewarded windows compared to 57 rewarded windows) due to the relatively low success rate in the pool task. Obtaining more samples is essential to discern an evident overall pattern. These two metrics are calculated for each participant individually and then averaged across participants.

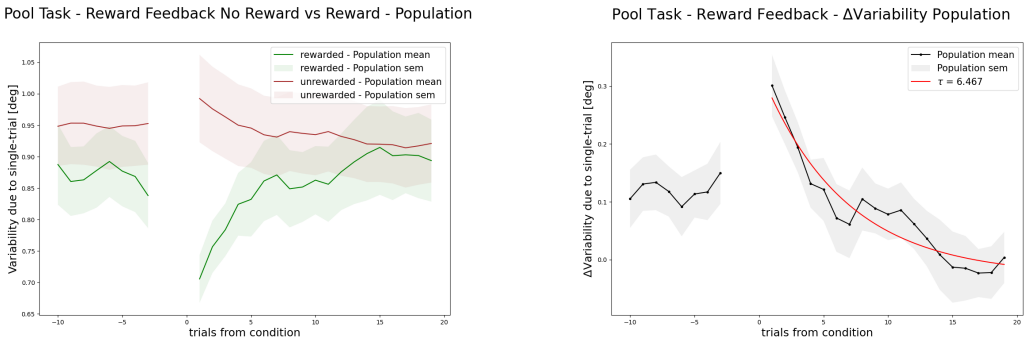


Figure 4.4.: Intertrial Variability following a conditioned trial average on the population

This time, a distinct pattern emerges. We observe an increase in variability following an unrewarded trial and a decrease after a rewarded trial. Additionally, the effect exhibits an exponential decrease with an average time constant of 6.4 trials.

To confirm that this effect can be attributed to the perceived binary feedback and its associated underlying mechanism, we conducted a parallel analysis on the pool task in the error feedback condition. In this scenario, where no reward is perceived by the participant, we observed the evolution of the variability conditioned on the success (or failure) of the condition trial. Importantly, in this condition, the participant is not aware of the outcome of the trial and only observes the sensory prediction error (i.e. the cue ball trajectory).
This time, we expect to observe no effect on the intertrial movement variability following the conditioned trial.
Specifically, we anticipate seeing neither an increase in variability following a failure nor a reduction in variability following a success. The modulation effect of the conditioned trial should be flat.

Figure 4.5.: Intertrial Variability following a conditioned trial average on the population in the pool task

A comparable analysis was conducted on the reaching task yielding similar results [A.1].

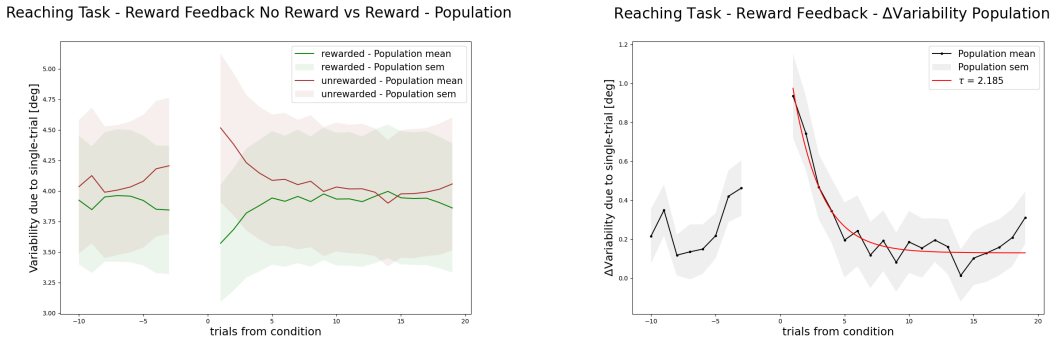


Figure 4.6.: Intertrial Variability following a conditioned trial average on the population in the reaching task

These findings suggest that participants use the reward information to modulate the exploration of motor commands when relying solely on binary reinforcement feedback. Furthermore, in the absence of reward feedback and when only sensory error feedback is available, intertrial movement variability seems not to be influenced by the task outcome. The disparity in the exploration behaviour suggests that the different adaptation trends previously observed [4.2] originate from different processing of those feedbacks.
This highlights the existence of two separate underlying mechanisms that result in distinct observed behaviours in each feedback condition.

The results align well with the theoretical expectations on the effect of reward on movement variability[81][95].
However, in contrast to prior findings in a traditional reaching task that indicated a stronger impact of the no-reward condition on variability and minimal effects in the reward condition [81], our observations in the pool task revealed a more pronounced modulation of variability in the rewarded condition.
These results suggest a modulation of variability regulated by the reward rate [95]. In a typical reaching task, the success rate is usually high due to the task's simplicity. Therefore, an unrewarded trial introduces a significant change in the reward rate, modulating variability more strongly. In contrast, in our more complex pool task with a relatively lower success rate, a rewarded trial introduces more substantial changes, leading to a more pronounced modulation of variability.

Chapter 5

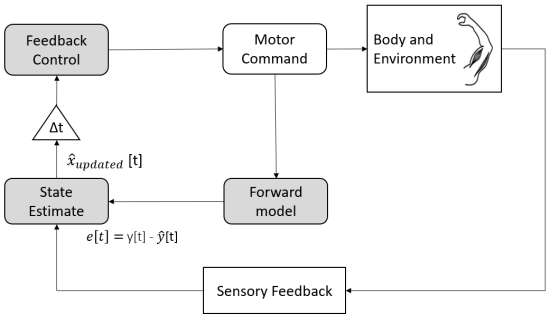
Model Architecture

While it is crucial for the computational models of motor adaptation to capture the observed behaviour, the objective of such models is to provide a potential explanation of the underlying mechanisms relating to each of the feedback. Therefore, we propose simple models robust to generalisation over different tasks, that require minimal assumptions and offer a clear explanation for the observed behaviour and fundamental functions of the nervous system. Those models are grounded in the previous findings in the field of sensorimotor adaptation [3].

To model the trial-to-trial adaptation of participants to a visuomotor perturbation in the realistic pool task and the reaching task, two different models are introduced, namely the error-based and the reward-based model.
Those two models, similarly to the participant, process those distinct feedbacks differently to generate an adaptation behaviour.

5.1. Error-based Model

Drawing from established literature that highlights the role of the Cerebellum as an internal forward model continuously revised through the integration of sensory prediction error [3.1], we have defined a perception-action loop to suit our particular tasks. This model consists of an internal forward model, constructing an internal representation of the world and predicting the expected sensory feedback, along with a controller responsible for selecting the most appropriate motor command based on these predictions.
The body and the world are treated as an unknown coupled dynamical system. When



considering the trial-to-trial angle adaptation in our tasks, several elements can be simplified.
Firstly, the motor outcome, representing the cue ball (or hand) movements, is expressed as a point mass. The movement on a given trial is defined by the angle of the cue ball (hand) trajectory from the initial position to the endpoint of that trial, denoted as h_n (see refdata analysis for a detailed definition of the angle).
Then, the neural signals defining the motor command u_n are believed to encode the limbs movements[51][52]. We define the high-level motor command u_n as the intended reach angle at time n. The motor outcome following a reaching movement can be expressed as the motor commands with some execution noise, where $h_n = u_n + \epsilon_n^m$.
Additionally, our interaction with the environment essentially involves visually observing the outcome, which can be perturbed by a rotation p_n. We consider this visual feedback as $y_n = h_n + p_n$.
This definition of the variables is consistent with typical models in prediction error models of sensorimotor adaptation [37][99][11].

By being repeatedly exposed to perturbation, participants build the prior knowledge that perturbations are correlated from trial to trial and subject to noise. This can be expressed as[11]:

$$\hat{p}_n = a_p \hat{p}_{n-1} + \epsilon_n^p$$

Where the coefficient a_p determines the strength of the correlation.
Subsequently, they generate predictions of the motor outcome $\hat{h}_n$ based on an efference copy of the motor command, an estimate of the system's dynamics (i.e. perturbation)$\hat{p}_n$, and some noise ϵ_n^h that contributes to the planning variability:

$$\hat{h}_n = \hat{p}_n + u_n + \epsilon_n^h.$$

The planning noise ϵ^x encompasses noise on the estimation of the perturbation and the estimation of the motor outcome.
This uncertainty estimate is equivalent to the process noise in the Kalman filter [3.1.3].
As the participant is unaware of the perturbation, the expected feedback estimated by the forward model is a noisy observation of the estimated motor outcome $\hat{y}_n = \hat{h}_n + \epsilon_n^m$.
The primary source of this noise originates from motor execution but is also influenced by noise in the sensory receptors. This noise source is the equivalent of the observation noise in the Kalman filter.

As we have standardised our angles relative to the task goal [4], our error model's objective is to adjust the motor command so that the observed feedback y_n corresponds to a 0° angle.

The system within our perception-action loop represents the underlying dynamics intervening between the motor command and the perceived sensory feedback. It involves two key equations: the noisy transformation of motor commands u_n into motor outcomes h_n, and then the transformation of the motor outcome h_n by the environment giving the observed feedback y_n. Therefore, akin to the approach of Shadmehr & al [11], we define the internal state x_n, which describes the underlying dynamics of our system, comprising an estimate of the hand position $\hat{h}_n$ and an estimate of the perturbation $\hat{p}_n$.

$$x_n = \begin{bmatrix} \hat{p}_n \\ \hat{h}_n \end{bmatrix} \tag{5.1}$$

Finally, our system is assumed to be time-invariant (i.e. the learner assumes that the same motor commands yield always the same sensory consequence). Therefore, the state transition, control input and observation matrices of our State Space Model are time-invariant.
We can then express the forward model as the following State Space Model:

$$\begin{aligned} &x_{n+1} = Ax_n + Bu_n + \epsilon_n^x, \\ &\epsilon_n^x = \begin{bmatrix} \epsilon_n^p \\ \epsilon_n^h \end{bmatrix} \sim \mathcal{N}(0, W_n), \quad W_n = \begin{bmatrix} \sigma_p^2 & 0 \\ 0 & \sigma_h^2 \end{bmatrix} \\ &\hat{y}_n = Cxn + \epsilon_n^m, \quad \epsilon_n^m \sim \mathcal{N}(0, V_n), \quad V_n = \begin{bmatrix} \sigma_m^2 \end{bmatrix} \\ &A = \begin{bmatrix} a_p & 0 \\ 1 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 1 \end{bmatrix} C = \begin{bmatrix} 0 & 1 \end{bmatrix} \end{aligned} \tag{5.2}$$

The presence of noise at various levels complicates the process of updating the forward model when new sensory data becomes available. In line with previous research [3.1], we employ a Kalman filter for updating the forward model.
The Kalman filter provides a solution for this by allowing the optimal integration of incoming sensory information [3.1.3].
The model adapts to the perturbation by iteratively updating the forward model based on the observed sensory feedback.

$$\begin{aligned} e_n &= y_n - \hat{y}_n \\ \hat{x}_{updated,n} &= A\hat{x}_n + Bu_n + Ke_n \end{aligned} \tag{5.3}$$

see [3.4] for the calculation of the Kalman gain K.
The two forms of uncertainty (i.e. planning and execution) have distinct influences on the adaptation process[100]. Planning noise contributes to the improvement of adaptation,

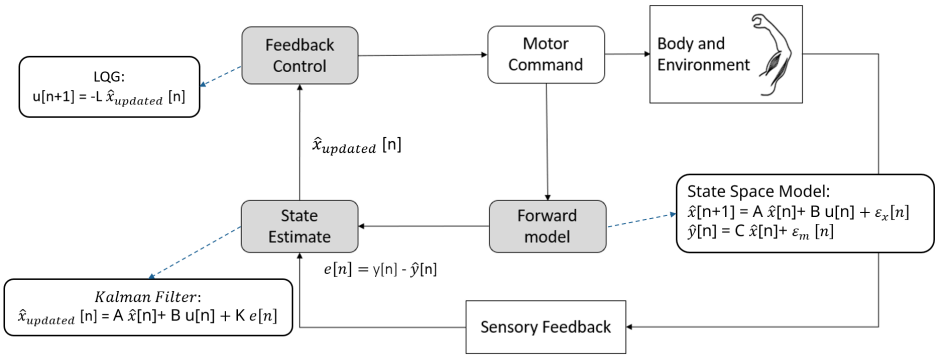
whereas uncertainty in the feedback estimate diminishes the adaptation rate, reflecting reduced confidence in the sensory feedback information.

Finally, the optimal motor command u_n given the updated state estimate of the system's dynamics can be computed using a Linear Quadratic Regulator akin to [3.1.4].

$$\begin{aligned} u_{n+1} &= -L\hat{x}_{updated,n} \\ \mathcal{J} &= \mathbb{E}[\sum_{i=0}^{N} x_i^T Q x_i + u_i^T R u_i] \\ Q &= 0.9 \\ R &= 0.1 \end{aligned} \tag{5.4}$$

see [3.1.4] for computation of the feedback gain L.

Essentially, the motor command is determined by revising the estimation of the system's dynamics as new sensory feedback arrives, and selecting an optimal command based on that revised estimate to achieve the desired goal at a minimum cost $\mathcal{J}$.
The general framework of this model has been introduced in [64] and its biological plausibility was discussed in [68]. Furthermore, the update of the forward model by the sensory prediction error relying on a Kalman filter has been successfully applied in sensory-motor adaptation [37][11].



5.2. Reward-based Model

Building on the current beliefs that the Basal Ganglia is involved in a form of reinforcement learning [3.3], we define a motor adaptation model that learns an estimation of the successful reach angle x_n based on binary sensory feedback of the motor consequences y_n and incorporates exploration noise.

We categorise three types of noise in our model:

- ϵ^m representing noise originating from motor command execution
- ϵ^p indicating noise at the planning stage of the movement.
- ϵ^e representing exploration noise that is actively regulated by the learner

In line with previous findings[58][59], we assume that motor variability follows a Gaussian distribution. Additionally, we assume that the learner is aware of the planning and exploration noise but ignorant of motor execution noise.

The intended reach angle (i.e. the mean of the distribution) is updated in response to binary sensory feedback and considering various sources of uncertainty. Two strategies have been proposed in the literature to address this [3.3].

The first approach suggests adjusting the optimal reach angle estimation based on the outcome of the last planned reach angle [3.7][3.8], or the last motor outcome 3.9 if the motor outcome was successful.
The second approach proposes to use a delta learning rule to update the estimate of the optimal reach angle after each trial, irrespective of the outcome [3.6] [3.10]. This delta learning rule computes the reward prediction error on the given trial, based on a reward prediction and the perceived reward. This method is grounded in the concept of reward prediction error and aligns with the reward system observed in the basal ganglia [3.3]. Furthermore, the motor variability decreases as a task is mastered, suggesting that both the regulated noise and the correction of the optimal reach angle estimation decrease as the learner observes regular success [4]. This leads us to adopt the second approach.

In reinforcement learning, estimating the value of an action in a continuous action space is typically addressed using the actor-critic architecture.
In this framework, the optimal policy is estimated by an actor responsible for the action selection, and the continuous value function is approximated by the critic, evaluating the value of an action.

In our task, consecutive actions exhibit correlations in the motor space[91]. Hence, a good estimate of the value of an action can be derived from recent performance history (i.e. reward rate), which has been linked to fluctuations in dopamine levels[101].

In line with [3.10], we define the value of an action as the reward rate, which can be computed as an exponentially weighted sum of past trials. Since the learner does not store the complete trial history in memory, an iterative method is employed to update the critic's value in an alternative formulation.
Additionally, the reward prediction error δ_n in our one timestep horizon problem, where consecutive trials are correlated, can be defined as the difference between the observed

reward and the predicted reward rate.

$$\begin{aligned} \bar{r}_{n+1} &= \bar{r}_n + \alpha_r \delta_n \\ \delta_n &= r_n - \bar{r}_n \end{aligned} \tag{5.5}$$

The actor is then calculated using a temporal difference learning rule [77], akin to the approach in [11][95]. In our model, the learner is aware of the exploration noise and therefore updates the intended angle accordingly.

$$x_{n+1} = x_n + \alpha_x \delta_n \epsilon_n^e \tag{5.6}$$

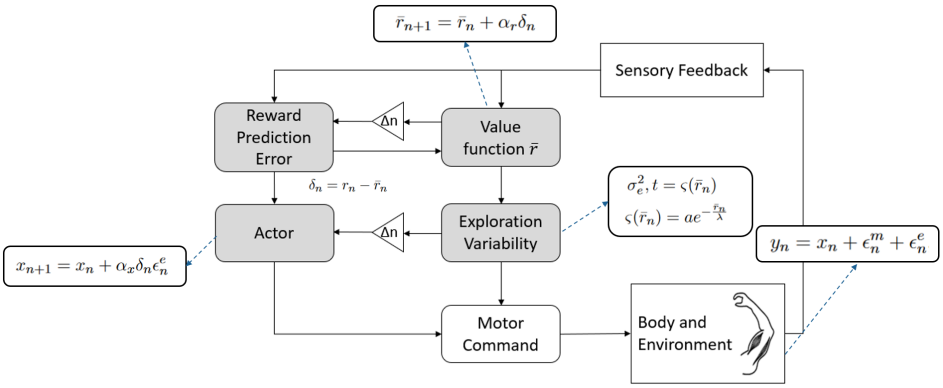
The observed angle (i.e. motor consequence) is drawn from a Gaussian distribution with mean x_n (i.e. intended angle), and variance comprising the three noise sources:

$$\begin{aligned} &y_n = x_n + \epsilon_n^m + \epsilon_n^p + \epsilon_n^e, \\ &\epsilon_n^m \sim \mathcal{N}(0, \sigma_m^2), \quad \epsilon_n^p \sim \mathcal{N}(0, \sigma_p^2), \quad \epsilon_n^e \sim \mathcal{N}(0, \sigma_e^2) \end{aligned} \tag{5.7}$$

Drawing from the insights provided in [95] [81], and the analysis presented in [4.4], we define the regulation of the exploration variability as an exponentially decreasing function of the reward rate.

$$\begin{aligned} &\sigma_e^2, t = \varsigma(\bar{r}_n) \\ &\varsigma(\bar{r}_n) = a e^{-\frac{\bar{r}_n}{\lambda}} \end{aligned} \tag{5.8}$$

Where a is the maximum amount of regulated movement variability, and λ is the time constant defining the exponential decay of the variability with the increasing reward rate(see [3.4]).



Chapter 6

Results

Observing a parallel trend between a model and experimental data isn't sufficient to establish that the model comprehensively explains the observed phenomena. However, a simple model constructed by introducing experimentally verified assumptions, and capturing the patterns of the experimental data, allows a clear interpretation of the underlying mechanisms.

To assess the capability of our models to explain the underlying mechanisms, we examined their ability to qualitatively capture the adaptation behaviour. To achieve this, we fitted the error model and reward model defined earlier to each participant and simulated the two feedback conditions in an adaptation session.
As previously mentioned, repeated attempts to execute the same movement invariably result in different outcomes [3]. Noise plays a crucial role in motor adaptation and is actively shaping the observed behaviour. Despite our models incorporating the different sources of motor variability, the limited number of data per participant (i.e. one adaptation session) complicates the comparison of the resulting outcomes.
Hence, we do not compare the exhibited behaviour at an individual level, aside from demonstrating the ability to display the wide range of behaviours observed in different participants (in terms of adaptation rate and noise level), particularly in the reward case [A.4].

On a broader scale, when examining the participant population, the influence of random noise is mitigated, allowing for a more meaningful comparison.
On both the reaching and the pool task, the reward-based and the error-based model visually showcased the ability to capture the adaptation to the perturbation displayed by the participants in the respective feedback conditions [6.1][6.3].

A more comprehensive evaluation of model fit is carried out using the following metrics:

- The Root Mean Squared Error (RMSE) provides an estimate of the typical deviation between the angles predicted by our model and those observed by the participants. It is a commonly used metric in regression analysis. Smaller (positive) RMSE values indicate a better fit.
- The coefficient of determination (R^2) assesses how well a model explains the variance in the observed variable. It quantifies the proportion of the variability in the data that can be predicted by the model. This metric is defined between 0 and 1, and a higher R^2 value indicates a better fit.
- Finally, we defined a Weighted Mean Squared Error (WMSE) to emphasise the transient phases, that is the early adaptation phase and the early washout phase.

6.1. Error-Based Model

The free parameters θ of the error model [5.1] are the state learning rate parameter:

$$a_p$$

and the noise parameters:

$$\sigma_p^2, \sigma_h^2, \sigma_m^2$$

The optimal parameters for each participant were found by maximising the log-likelihood of the observed data over a grid of parameters [3.5].

Once the parameters fitted to each participant, we ran one model simulation per subject and compared the average resulting behaviour over the population between the experimental data and our model's prediction.
In the pool task, we note the following distribution of parameters:

Error Model Parameters - Pool task								
Parameters	a_p		σ_p^2		σ_h^2		σ_m^2	
Value	0.749	$\pm$	2.881	$\pm$	2.881	$\pm$	0.301	$\pm$
	0.268		0.283		0.283		0.309	

To further mitigate the impact of noise, we assess the average angle, the success rate, and the movement variability by averaging the metrics over blocks of 25 trials similarly to [A.2].
It is worth noting that, in the pool task, the participants were provided with full feedback during the baseline and washout phases.

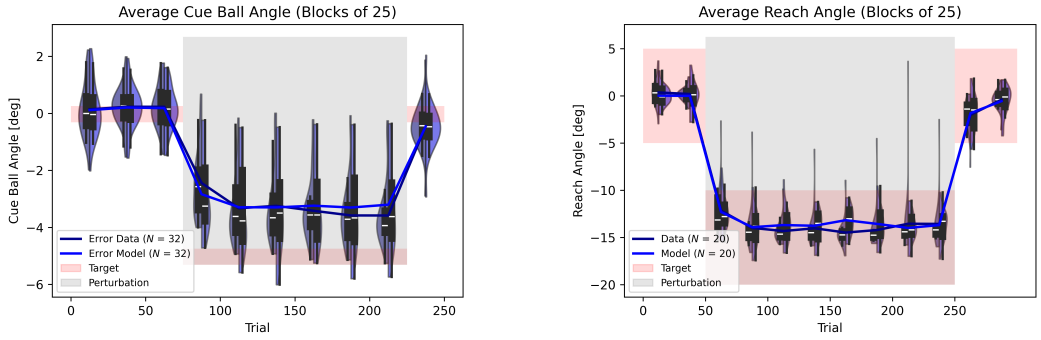


Figure 6.1.: Comparison of the block averaged angle of the Error Model and the participants on the pool task (left) and reaching task (right) in the error condition.

The directional error is similar between the model and the data in both tasks. Furthermore, the double-exponential learning curve characterising adaptation relying on sensory prediction error [37] is well captured.
A block analysis of the success rate and the variability was then conducted, akin to [A.2]. The variability is computed as the standard deviation, corrected with respect to the linear regression in the block.

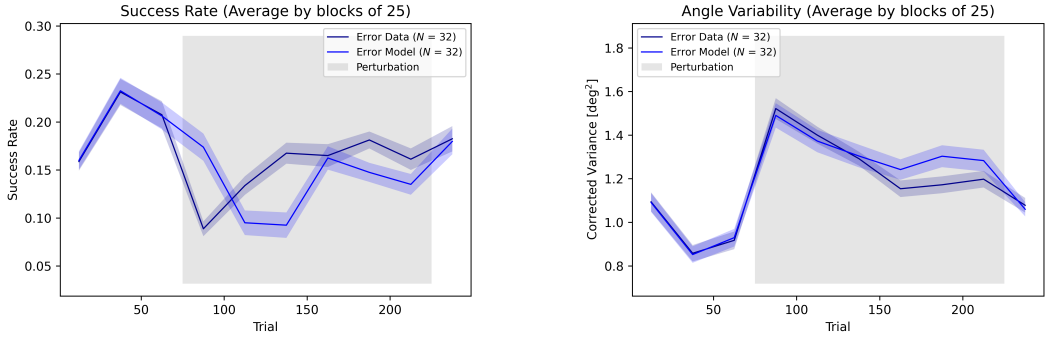


Figure 6.2.: Success Rate (left) and variability (right) per block over the population in the Reaching task

Remarkably, our model successfully replicates the immediate increase in variability during the initial phase of the perturbation when an error is observed and exploration is needed, followed by a gradual decline during the adaptation process as the sensory remapping of the forward model occurs.
The success rate decreases upon the introduction of the perturbation, followed by a gradual increase during the adaptation process.

We observe the following performance metrics after fitting the model to each of the participants.

Error Model Performance - Pool task			
Metrics	$RMSE$	R^2	$WMSE$
Value over Population	0.868	0.9089	0.057
Value over Subject	2.248 ± 0.440	0.391 ± 0.134	0.122 ± 0.031

The low RMSE and WMSE indicate minimal deviations between the model's predictions and the data. The high R-squared (R^2) value, close to 1, indicates that the model effectively accounts for a substantial portion of the variability present in the data.
The metrics show a considerable decrease in performance when proceeding to individual comparison in contrast with comparison over the population average.
As mentioned previously [3], this difference is attributed to the substantial variability in both the planning and execution of movements [3], and considering that the endpoints of movements are normally distributed [58], the resulting behaviour is closer to the expected value when averaging over a large number.

6.2. Reward-based Model

Due to the definition of the dynamic reward zone during the perturbation phase, the reward depends on the last ten rewarded angles, and an analytic fit is not tractable. Therefore, akin to [82] we used a particle filter to estimate the likelihood of the observed data given the model's parameters.

The particle filter involves generating N ($N = 500$ in our case) particles, each representing a potential value for the state variable being estimated (e.g., angle in our case). Initially, each particle has an equal weight, indicating its likelihood to match the current state of the process we are modelling (i.e. observed data). These particles are then processed through the model, producing a new state prediction. The weight of each particle is updated based on the probability of observing the measurement given that new state prediction. Subsequently, particles are resampled with replacement according to their weight. Repeating this process throughout the adaptation session leads to convergence toward particles that are most likely to explain the data given the model's parameters. The likelihood of the observed data given the model's parameters is then computed as the sum of the weights across all particles. Optimal parameters are then found by maximising this likelihood over a grid of parameter values.

The free parameters θ of the reward model ([5.5] to [5.8]) are the learning rate parameters: α_x, α_r
the fixed noise parameters: σ_m^2, σ^p
and the regulated noise parameters: a, λ

Similarly to the error condition, we conducted one model simulation per subject with parameters fitted accordingly, comparing the average resulting behaviour between the experiments and our model's predictions.

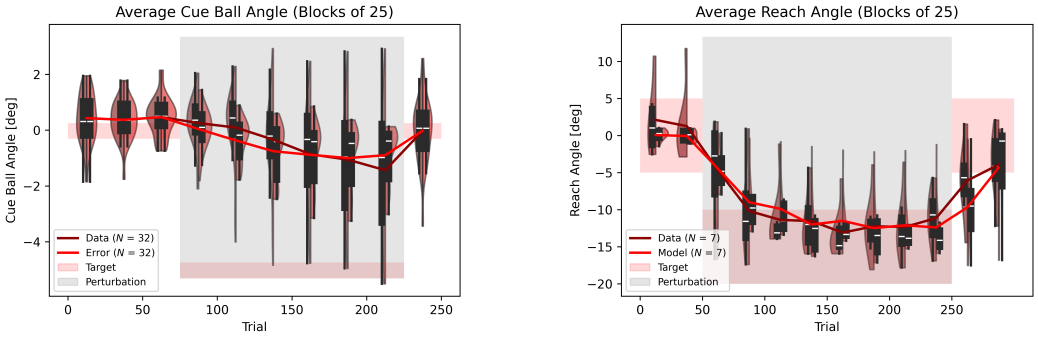


Figure 6.3.: Comparison of the block averaged angle of the Reward Model and the participants on the pool task (left) and reaching task (right) in the reward condition.

We note that the model displays between subjects variability and generally replicates the suboptimal learning trend observed in the participants that can arise as a result when greater emphasis is placed on locally optimal information rather than globally optimal [83].
The model demonstrates a more pronounced exponential learning curve than the participants. This difference is likely due to the parameter tuning process. Using a grid search with seven parameters introduces computational complexity, and further refinement may help bridge the gap and align the model with the average participant's behaviour. We have struck a balance between time complexity and accuracy to showcase the model's qualitative capturing capabilities.

Several examples of individual subject-wise comparisons illustrate that our model effectively captures the wide range of behaviour observed in the reward condition and displays within-subjects variability [A.4].
The parameters of the reward models in the pool task are distributed as follows:

Reward Model Parameters - Pool task						
Parameters	α_x	α_r	σ_m^2	σ_p^2	λ	a
Value	0.113 ± 0.056	0.890 ± 0.097	0.128 ± 0.076	0.808 ± 0.054	0.946 ± 0.593	1.468 ± 0.754

We then conducted a block analysis to examine the motor variability and success rate patterns of the reward model in comparison to the data.

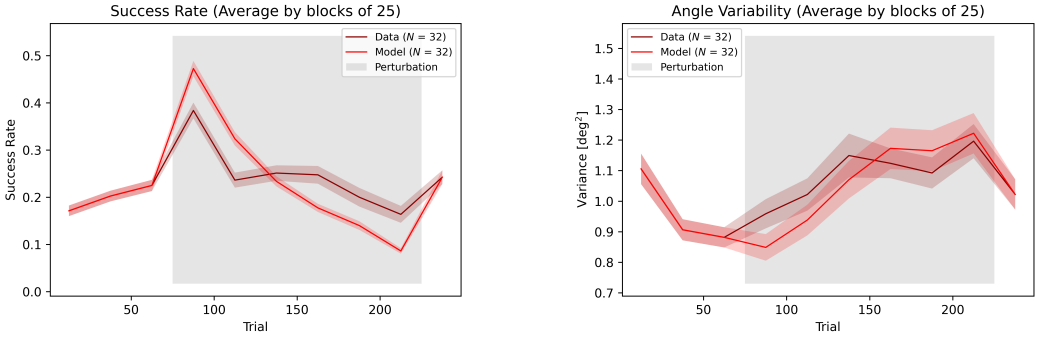


Figure 6.4.: Success Rate (left) and variability (right) per block over the population in the Pool task

The reward model effectively reproduces the observed patterns in the evolution of the success rate and variability during the adaptation phase.
Participants exhibit a relatively low success rate on the baseline when they have full feedback, and their decreasing variability suggests ongoing learning of the optimal angle. Initially, the success rate remains relatively high in the perturbation phase, due to the dynamic reward zone definition. However, the reward zone gradually decreases and getting rewarded becomes more challenging. This is reflected both in the decreasing success rate and in the increased variability throughout the adaptation phase.

We observe the following performance metrics after fitting the model to each of the participants.

Reward Model Performance - Pool task			
Metrics	$RMSE$	R^2	$WMSE$
Value over Population	0.693	0.494	0.082
Value over Subject	1.87 ± 0.802	0.051 ± 0.007	0.118 ± 0.057

The low RMSE and WMSE indicate small deviations between the model's predictions and the data, whereas the relatively high R-squared (R^2) value (closer to 1 the better),

indicates that the model explains a substantial amount of the variability present in the data.
Again, we see an important decrease in performance when comparing the average behaviour of the model to the average behaviour of the participants and the individual Subject-wise comparisons.

Overall, we observe the following adaptation behaviour emerging from the models compared with the experimental observations.

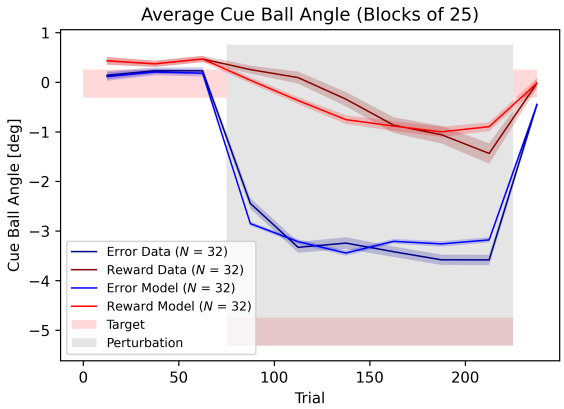


Figure 6.5.: Comparison of the average angle between the Error and Reward models and the participant's data in the pool task

Finally, an examination of the intertrial variability similar to [4.4] is conducted in the model simulations.
For the reward-based model simulations, as expected given the model's design, we observe an exponentially decreasing impact on variability, explaining the patterns observed in the experimental data under the reward condition [4.4].

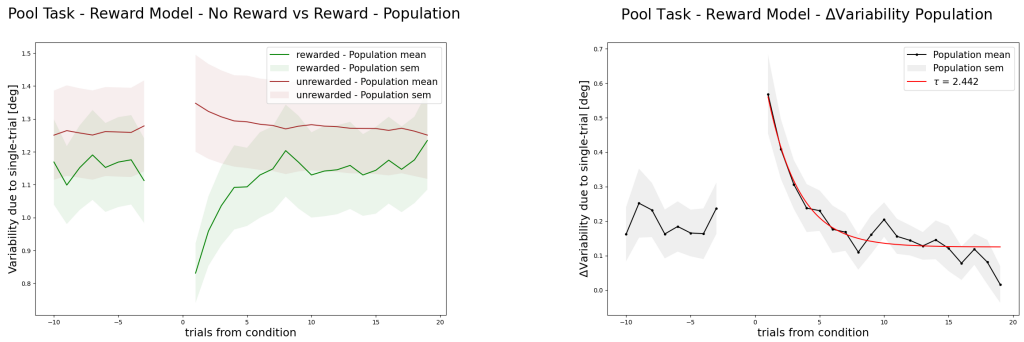


Figure 6.6.: Reward model variability following a conditioned trial average on the simulated population in the pool task

In contrast, the analysis of the error model reveals no distinctive modulation of variability associated with the task outcome, consistent with the data observed in the error condition [4.5].

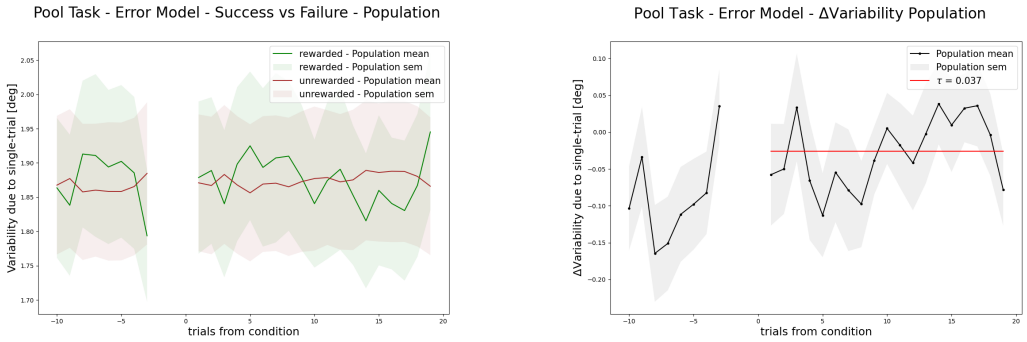


Figure 6.7.: Error model variability following a conditioned trial average on the simulated population in the pool task

The reward model successfully captures and explains the reward modulation of the inter-trial variability exhibited in the reward condition of both the pool task and the reaching task.

Chapter 7

Conclusion and Future Avenues

Our study aims to explore the coexistence of two distinct mechanisms thought to take place in the cerebellum and basal ganglia [19].
This concept is grounded in observations of patients with impairments in those specific structures. The deficits exhibited by these patients in motor tasks led researchers to the conclusion that these brain regions are crucial for motor learning[17][80][18]. Furthermore, it was observed that despite showing motor deficits resulting from the impairment of either of those areas, learning remained possible[82][81].

Based on the anatomical characteristics of these brain regions, it has been proposed [19]that the cerebellum constructs an internal model of the world to predict the consequences of motor commands. The disparities between predictions and observed outcomes subsequently enable the adjustment of behaviour. Whereas the basal ganglia rely on binary reinforcement feedback to adjust the movement.
It has been demonstrated that implicit movement adaptation can be achieved through reliance on either sensory prediction error or reward feedback[11]. Additionally, when perceiving both, participants typically show a greater reliance on sensory feedback error. However, when confronted with the multiple degrees of freedom and variables in a real-world task, such as playing pool, opposing trends in neural biomarkers thought to signal the learning related to specific feedback suggest that participants adopt different strategies[26].

The nature of the information provided by these two feedbacks indicates the use of different underlying mechanisms.
To further explore this, the real-world pool task was adapted to provide specific feedback on the examination of each underlying mechanism individually. To establish a baseline for our findings, a traditional visuomotor rotation hand-reaching task provided specific visual feedback was conducted in the laboratory.

This project conducts a comprehensive analysis of the experiments, with a primary focus on the reward condition, as the error condition is already well-established in the literature. Following a thorough review of the motor adaptation field and an analysis of the observed behaviours in the experiments, we defined two simple adaptation models adapted to our experiments. These models make minimal assumptions, are supported by anatomical evidence related to the cerebellum and basal ganglia and include experimentally verified concepts to exploit the information conveyed by the two feedbacks differently. They allow a clear interpretation of the underlying high-level processes and fundamental function of the nervous system.

Fitting the parameters of these models to the participants who received error feedback in the reaching and pool tasks showed to be consistent with the prevailing theory of learning based on sensory prediction error using an internal forward model that is updated through Bayesian integration.
In the context of reward feedback, the model captured the overall adaptation pattern in terms of average behaviour, success rate and variability in both tasks. Additionally, the model exhibited characteristic features of reward-based adaptation, including suboptimality [83] and reward-dependent variability modulation[81].
This confirms that the learner actively regulates variability in his movements based on the performance history to explore the motor space and maximise the reward in a real-world task.

In conclusion, the high-level models proposed confirmed several hallmarks of sensorimotor adaptation in a real-world task. They require minimal assumptions and allow for explainability of the underlying processes, providing a solid foundation for further research into sensorimotor adaptation.

Future steps involve understanding how these mechanisms are combined. A simple weighting between the commands generated by each adaptation model could allow each participant to adopt a different adaptation strategy. Alternatively, a more complex combination involving meta-learning [102] holds great promise.

These models present promising avenues for enhancing our comprehension of the functional roles played by the cerebellum and basal ganglia in sensorimotor adaptation. This, in turn, contributes to a better understanding of neurodegenerative diseases affecting these structures, such as cerebellar stroke, cerebellar ataxia, or Parkinson's disease.
Specifically, the real-world pool task setup could be adapted to detect the absence of the reward modulation of movement variability in patients suffering from Parkinson's disease. Furthermore, personalised therapy could be developed by intelligently manipulating visual feedback for patients to maximise its effectiveness as suggested in[103].
Finally, this research contributes to a deeper understanding of the overall functioning of the human brain.

Bibliography

[1] "github repository of the project," https://github.com/dario-bolli/motor_learning_model.

[2] N. Bernstein, *The Co-ordination and Regulation of Movements*. Pergamon Press, 1967.

[3] A. A. Faisal, L. P. J. Selen, and D. M. Wolpert, "Noise in the nervous system," 2008. [Online]. Available: https://www.nature.com/articles/nrn2258

[4] J. W. Krakauer and P. Mazzoni, "Human sensorimotor learning: adaptation, skill, and beyond," 2011. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/21764294/

[5] D. Wolpert, J. Diedrichsen, and J. R. Flanagan, "Principles of sensorimotor learning," 2011. [Online]. Available: https://www.nature.com/articles/nrn3112

[6] J. W. Krakauer, A. M. Hadjiosif, J. Xu, A. L. Wong, and A. M. Haith, "Motor learning," 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/cphy.c170043

[7] R. Shadmehr and F. Mussa-Ivaldi, "Adaptive representation of dynamics during learning of a motor task," 1994. [Online]. Available: https://www.jneurosci.org/content/14/5/3208

[8] J. W. Krakauer, "Motor learning and consolidation: the case of visuomotor rotation," 2009. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2672910/

[9] J. W. Krakauer, Z. M. Pine, M. F. Ghilardi, and C. Ghez, "Learning of visuomotor transformations for vectorial planning of reaching trajectories," 2000. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/11102502/

[10] J. A. Taylor, J. W. Krakauer, and R. B. Ivry, "Explicit and implicit contributions to learning in a sensorimotor adaptation task," 2014. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/24553942/

[11] J. Izawa and R. Shadmehr, "Learning from sensory and reward prediction errors during motor adaptation," 2011. [Online]. Available: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002012

[12] P. Mazzoni and J. W. Krakauer, "An implicit plan overrides an explicit strategy during visuomotor adaptation," 2006. [Online]. Available: https://www.jneurosci.org/content/26/14/3642

[13] R. Shadmehr, M. A. Smith, and J. W. Krakauer, "Error correction, sensory prediction, and adaptation in motor control," 2010. [Online]. Available: https://www.annualreviews.org/doi/10.1146/annurev-neuro-060909-153135

[14] D. Silver, S. Singh, D. Precup, and R. S. Sutton, "Reward is enough," 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0004370221000862

[15] N. M. van Mastrigt, J. S. Tsay, T. Wang, G. Avraham, S. J. Abram, K. van der Kooij, J. B. J. Smeets, and R. B. Ivry, "Implicit reward-based motor learning," 2023. [Online]. Available: https://link.springer.com/article/10.1007/s00221-023-06683-w#Sec2

[16] S. Ueharam, F. Mawase, and P. Celnik, "Learning similar actions by reinforcement or sensory-prediction errors rely on distinct physiological mechanisms," 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6887949/

[17] J. Izawa, S. E. Criscimagna-Hemminger, and R. Shadmehr, "Cerebellar contributions to reach adaptation and learning sensory consequences of action," 2012. [Online]. Available: https://www.jneurosci.org/content/32/12/4230

[18] H. Krebs, N. Hogan, W. Hening, S. Adamovich, and H. Poizner, "Procedural motor learning in parkinson's disease," 2001. [Online]. Available: https://link.springer.com/article/10.1007/s002210100871

[19] K. Doya, "Complementary roles of basal ganglia and cerebellum in learning and motor control," 2000. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959438800001537?via%3Dihub

[20] S. Haar, C. M. van Assel, and A. Faisal, "Motor learning in real-world pool billiards," 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/33208785/

[21] A. K. Roopun, S. J. Middleton, M. O. Cunningham, F. E. N. LeBeau, A. Bibbig, M. A. Whittington, and R. D. Traub, "A beta2-frequency (20-30 hz) oscillation in nonsynaptic networks of somatosensory cortex," 2006. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/17030821/

[22] H. Tan, N. Jenkinson, and P. Brown, "Dynamic neural correlates of motor error monitoring and adaptation during trial-to-trial learning," 2014. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/24741058/

[23] F. Torrecillos, J. Alayrangues, B. E. Kilavik, and N. Malfait, "Distinct modulations in sensorimotor postmovement and foreperiod -band activities related to error salience processing and sensorimotor adaptation," 2015. [Online]. Available: https://www.jneurosci.org/content/35/37/12753

[24] C. Kranczioch, S. Athanassiou, S. Shen, G. Gao, and A. Sterr, "Short term learning of a visually guided power-grip task is associated with dynamic changes in eeg oscillatory activity," 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1388245708001569

[25] S. Haar and A. Faisal, "Neural biomarkers of multiple motor-learning mechanisms in a real-world task," 2020. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnhum.2020.00354/full

[26] ——, "Brain activity reveals multiple motor-learning mechanisms in a real-world task," 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32982707/

[27] S. Haar, G. Sundar, and A. Faisal, "Embodied virtual reality for the study of real-world motor learning," 2020. [Online]. Available: https://www.biorxiv.org/content/10.1101/2020.03.19.998476v1

[28] F. Nardi, M. Ziman, S. Haar, and A. Faisal, "Isolating motor learning mechanisms in embodied virtual reality," 2022. [Online]. Available: https://www.researchgate.net/publication/362707312_Isolating_Motor_Learning_Mechanisms_in_Embodied_Virtual_Reality

[29] H. Li and S. Haar, "Neural signature of different learning mechanisms in motor learning," 2022.

[30] A. M. Haith and J. W. Krakauer, "Model-based and model-free mechanisms of human motor learning," 2013. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3570165/

[31] A. Galvan, A. Devergnas, and T. Wichmann, "Alterations in neuronal activity in basal ganglia-thalamocortical circuits in the parkinsonian state," 2015. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4318426/

[32] P. N. Tobler, C. D. Fiorillo, and W. Schultz, "Adaptive coding of reward value by dopamine neurons," 2005. [Online]. Available: https://www.science.org/doi/10.1126/science.1105370

[33] J. C. Houk, J. L. Adams, and A. G. Barto, "A model of how the basal ganglia generate and use neural signals that predict reinforcement," 1994. [Online]. Available: https://ieeexplore.ieee.org/document/6287670

[34] J. P. Scholz and G. Schöner, "The uncontrolled manifold concept: identifying control variables for a functional task," 1999. [Online]. Available: https://link.springer.com/article/10.1007/s002210050738

[35] F. Mussa-Ivaldi, "Do neurons in the motor cortex encode movement direction? an alternative hypothesis," 1988. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/3173781/

[36] T. Ikegami, J. R. Flanagan, and D. M. Wolpert, "Reach adaption to a visuomotor gain with terminal error feedback involves reinforcement learning," 2022. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0269297

[37] M. A. Smith, A. Ghazizadeh, and R. Shadmehr, "Interacting adaptive processes with different timescales underlie short-term motor learning," 2006. [Online]. Available: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0040179

[38] D. M. Wolpert, R. Miall, and M. Kawato, "Internal models in the cerebellum," 1998. [Online]. Available: https://www.cell.com/trends/cognitive-sciences/fulltext/S1364-6613(98)01221-2?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS1364661398012212%3Fshowall%3Dtrue

[39] R. J. van Beers, A. C. Sittig, and J. J. D. van der Gon, "Integration of proprioceptive and visual position-information: An experimentally supported model," 1999. [Online]. Available: https://journals.physiology.org/doi/full/10.1152/jn.1999.81.3.1355

[40] R. A. Scheidt, M. A. Conditt, E. L. Secco, and F. A. Mussa-Ivaldi, "Interaction of visual and proprioceptive feedback during adaptation of human reaching movements," 2005. [Online]. Available: https://journals.physiology.org/doi/full/10.1152/jn.00947.2004?rfr_dat=cr_pub++0pubmed&url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org

[41] M. Kawato and H. Gomi, "A computational model of four regions of the cerebellum based on feedback-error learning," 1992. [Online]. Available: https://link.springer.com/article/10.1007/BF00201431

[42] R. C. M. D M. Wolpert, "Forward models for physiological motor control," 1996. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/12662535/

[43] Y.-W. Tseng, J. Diedrichsen, J. W. Krakauer, R. Shadmehr, and A. J. Bastian, "Sensory prediction errors drive cerebellum-dependent adaptation of reaching," 2007. [Online]. Available: https://journals.physiology.org/doi/full/10.1152/jn.00266.2007?rfr_dat=cr_pub++0pubmed&url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org

[44] D. E. R. Michael B. Jordan, "Forward models: Supervised learning with a distal teacher," 1996. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1207/s15516709cog1603_1

[45] R. B. I. John Schlerf and J. Diedrichsen, "Encoding of sensory prediction errors in the human cerebellum," 2012. [Online]. Available: https://www.jneurosci.org/content/32/14/4913

[46] S.-J. Blakemore, C. D. Frith, and D. M. Wolpert, "Spatio-temporal prediction modulates the perception of self-produced stimuli," 1999. [Online]. Available: https://direct.mit.edu/jocn/article-abstract/11/5/551/3376/Spatio-Temporal-Prediction-Modulates-the?redirectedFrom=fulltext

[47] R. Shadmehr and J. W. Krakauer, "A computational neuroanatomy for motor control," 2008. [Online]. Available: https://link.springer.com/article/10.1007/s00221-008-1280-5

[48] K. A. Thoroughman and R. Shadmehr, "Electromyographic correlates of learning an internal model of reaching movements," 1999. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6783008/

[49] M. Kawato, "Internal models for motor control and trajectory planning," 1999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959438899000288?via%3Dihub

[50] H. Imamizu, S. Miyauchi, T. Tamada, Y. Sasaki, R. Takino, B. Pütz, T. Yoshioka, and M. Kawato, "Human cerebellar activity reflecting an acquired internal model of a new tool," 2000. [Online]. Available: https://www.nature.com/articles/35003194

[51] S. Pasalar, A. V. Roitman, W. K. Durfee, and T. J. Ebner, "Force field effects on cerebellar purkinje cell discharge with implications for internal models," 2006. [Online]. Available: https://www.nature.com/articles/nn1783

[52] A. V. Roitman, S. Pasalar, M. T. V. Johnson, and T. J. Ebner, "Position, direction of movement, and speed tuning of cerebellar purkinje cells during circular manual tracking in monkey," 2005. [Online]. Available: https://www.jneurosci.org/content/25/40/9244.long

[53] K. Körding and D. Wolpert, "Bayesian integration in sensorimotor learning," 2004. [Online]. Available: https://www.nature.com/articles/nature02169

[54] M. Berniker and K. Kording, "Estimating the sources of motor errors for adaptation and generalization," 2008. [Online]. Available: https://www.nature.com/articles/nn.2229

[55] A. Korenberg and Z. Ghahramani, "A bayesian view of motor adaptation," 2002. [Online]. Available: https://www.semanticscholar.org/paper/A-bayesian-view-of-motor-adaptation-Korenberg-Ghahramani/272fa3f644194dbc10d13607ad21fcda317c83cb

[56] A. Stocker and E. Simoncelli, "Noise characteristics and prior expectations in human visual speed perception," 2006. [Online]. Available: https://www.nature.com/articles/nn1669

[57] M. d'Acremont, W. Schultz, and P. Bossaerts, "The human brain encodes event frequencies while forming subjective beliefs," 2013. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4293915/

[58] J. Trommershäuser, S. Gepshtein, L. T. Maloney, M. S. Landy, and M. S. Banks, "Optimal compensation for changes in task-relevant movement variability," 2005. [Online]. Available: https://www.jneurosci.org/content/25/31/7169

[59] M. S. Landy, J. Trommershäuser, and N. D. Daw, "Dynamic estimation of task-relevant variance in movement under risk," 2012. [Online]. Available: https://www.jneurosci.org/content/32/37/12702

[60] D. M. Wolpert, Z. Ghahramani, and M. . Jordan, "An internal model for sensorimotor integration," 1995. [Online]. Available: https://www.science.org/doi/10.1126/science.7569931?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200pubmed

[61] D. M. Wolpert, "Computational approaches to motor control," 1997. [Online]. Available: https://www.cell.com/trends/cognitive-sciences/pdf/S1364-6613(97)01070-X.pdf?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS136466139701070X%3Fshowall%3Dtrue

[62] K. Wei and K. Körding, "Uncertainty of feedback and state estimation determines the speed of motor adaptation," 2010. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fncom.2010.00011/full

[63] M. Fujita, "Adaptive filter model of the cerebellum," 1982. [Online]. Available: https://link.springer.com/article/10.1007/BF00336192

[64] E. Todorov and M. I. Jordan, "Optimal feedback control as a theory of motor coordination," 2002. [Online]. Available: https://www.nature.com/articles/nn963

[65] M. Davis and R. Vinter, *Stochastic Modelling and Control*. Chapman and Hall, 1985.

[66] D. Liu and E. Todorov, "Evidence for the flexible sensorimotor strategies predicted by optimal feedback control," 2007. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/17728449/

[67] E. Todorov and M. I. Jordan, "A minimal intervention principle for coordinated movement," 2002. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2002/hash/8c5f6ecd29a0eb234459190ca51c16dd-Abstract.html

[68] S. H. Scott, "Optimal feedback control and the neural basis of volitional motor control," 2004. [Online]. Available: https://www.nature.com/articles/nrn1427

[69] J. Izawa, T. Rane, O. Donchin, and R. Shadmehr, "Motor learning as a process of reoptimization," 2008. [Online]. Available: https://www.jneurosci.org/content/28/11/2883

[70] "Bayes' Theorem," https://plato.stanford.edu/archives/spr2019/entries/bayes-theorem/, joyce, James (2003), The Stanford Encyclopedia of Philosophy (Spring 2019 ed.), Metaphysics Memoir on the Probability of the Causes of Events.

[71] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[72] "Roger labbe github," https://github.com/rlabbe.

[73] G. Goodwin and K. Sin, *Adaptive filtering, prediction and control.* Prentice-Hall, 1985.

[74] J. Burge, M. O. Ernst, and M. S. Banks, "The statistical determinants of adaptation rate in human reaching," 2008. [Online]. Available: https://jov.arvojournals.org//8/4/20/

[75] W. Schultz, "Predictive reward signal of dopamine neurons," 1998. [Online]. Available: https://journals.physiology.org/doi/full/10.1152/jn.1998.80.1.1

[76] W. Schultz, P. Dayan, and P. R. Montague, "A neural substrate of prediction and reward," 1997. [Online]. Available: https://www.science.org/doi/10.1126/science.275.5306.1593?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200pubmed

[77] R. E. Suri, "Td models of reward predictive responses in dopamine neurons," 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608002000461

[78] A. V. Kravitz, B. S. Freeze, P. R. L. Parker, K. Kay, M. T. Thwin, K. Deisseroth, and A. C. Kreitzer, "Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry," 2010. [Online]. Available: https://www.nature.com/articles/nature09159

[79] S. X. Luo and E. J. Huang, "Dopaminergic neurons and brain reward pathways," 2015. [Online]. Available: https://ajp.amjpathol.org/article/S0002-9440(15)00646-X/fulltext

[80] P. Mazzoni, A. Hristova, and J. W. Krakauer, "Why don't we move faster? parkinson's disease, movement vigor, and implicit motivation," 2007. [Online]. Available: https://www.jneurosci.org/content/27/27/7105

[81] S. E. Pekny, J. Izawa, and R. Shadmehr, "Reward-dependent modulation of movement variability," 2015. [Online]. Available: https://www.jneurosci.org/content/35/9/4015.long

[82] A. S. Therrien, D. M. Wolpert, and A. J. Bastian, "Effective reinforcement learning following cerebellar damage requires a balance between exploration and motor noise," 2015. [Online]. Available: https://academic.oup.com/brain/article/139/1/101/2468812?login=false

[83] J. G. A. Cashaback, C. K. Lao, D. J. Palidis, S. K. Coltman, H. R. McGregor, and P. L. Gribble, "The gradient of the reinforcement landscape influences sensorimotor learning," 2019. [Online]. Available: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006839

[84] S. Uehara, F. Mawase, and P. Celnik, "Learning similar actions by reinforcement or sensory-prediction errors rely on distinct physiological mechanisms," 2017. [Online]. Available: https://academic.oup.com/cercor/article/28/10/3478/4157545?login=false

[85] R. Shadmehr, "Generalization as a behavioral window to the neural mechanisms of learning internal models," 2004. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0167945704000235?via%3Dihub

[86] H. G. Wu, Y. R. Miyamoto, L. N. G. Castro, B. P. Ölveczky, and M. A. Smith, "Temporal structure of motor variability is dynamically regulated and predicts motor learning ability," 2014. [Online]. Available: https://www.nature.com/articles/nn.3616

[87] R. J. van Beers, P. Haggard, and D. M. Wolpert, "The role of execution noise in movement variability," 2004. [Online]. Available: https://journals.physiology.org/doi/full/10.1152/jn.00652.2003

[88] A. S. Therrien, D. M. Wolpert, and A. J. Bastian, "Increasing motor noise impairs reinforcement learning in healthy individuals," 2018. [Online]. Available: https://www.eneuro.org/content/5/3/ENEURO.0050-18.2018

[89] R. J. van Beers, E. Brenner, and J. B. J. Smeets, "Random walk of motor planning in task-irrelevant dimensions," 2013. [Online]. Available: https://journals.physiology.org/doi/full/10.1152/jn.00706.2012?rfr_dat=cr_pub++0pubmed&url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org

[90] M. M. Churchland, A. Afshar, and K. V. Shenoy, "A central source of movement variability," 2006. [Online]. Available: https://www.cell.com/neuron/fulltext/S0896-6273(06)00871-3?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0896627306008713%3Fshowall%3Dtrue

[91] R. J. van Beers, "Motor learning is optimally tuned to the properties of motor noise," 2009. [Online]. Available: https://www.cell.com/neuron/fulltext/S0896-6273(09)00516-9?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0896627309005169%3Fshowall%3Dtrue

[92] B. P. Ölveczky, A. S. Andalman, and M. S. Fee, "Vocal experimentation in the juvenile songbird requires a basal ganglia circuit," 2005. [Online]. Available: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0030153

[93] A. Leblois, B. J. Wendel, and D. J. Perkel, "Striatal dopamine modulates basal ganglia output and regulates social context-dependent behavioral variability through d1 receptors," 2010. [Online]. Available: https://www.jneurosci.org/content/30/16/5730

[94] N. M. van Mastrigt, J. B. J. Smeets, and K. van der Kooij, "Quantifying exploration in reward-based motor learning," 2020. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0226789

[95] A. K. Dhawale, Y. R. Miyamoto, M. A. Smith, and B. P. lveczky, "Reward-dependent modulation of movement variability," 2019. [Online]. Available: https://www.cell.com/current-biology/fulltext/S0960-9822(19)31102-9?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0960982219311029%3Fshowall%3Dtrue

[96] A. M. Roth, J. A. Calalo, R. Lokesh, S. R. Sullivan, S. Grill, J. J. Jeka, K. van der Kooij, M. J. Carter, and J. G. A. Cashaback, "Reinforcement-based processes actively regulate motor exploration along redundant solution manifolds," 2023. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/37848061/

[97] D. G. Tervo, M. Proskurin, M. Manakov, M. Kabra, A. Vollmer, K. Branson, and A. Y. Karpova, "Behavioral variability through stochastic choice and its gating by anterior cingulate cortex," 2014. [Online]. Available: https://www.cell.com/cell/fulltext/S0092-8674(14)01107-6?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0092867414011076%3Fshowall%3Dtrue

[98] S. T. Albert and R. Shadmehr, "Estimating properties of the fast and slow adaptive processes during sensorimotor adaptation," 2018. [Online]. Available: https://journals.physiology.org/doi/full/10.1152/jn.00197.2017?rfr_dat=cr_pub++0pubmed&url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org

[99] M. S. Fine and K. A. Thoroughman, "Trial-by-trial transformation of error into sensorimotor adaptation changes with environmental dynamics," 2007. [Online]. Available: https://journals.physiology.org/doi/full/10.1152/jn.00196.2007?rfr_dat=cr_pub++0pubmed&url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org

[100] J. Burge, M. O. Ernst, and M. S. Banks, "The statistical determinants of adaptation rate in human reaching," 2008. [Online]. Available: https://jov.arvojournals.org/article.aspx?articleid=2122298

[101] A. A. Hamid, J. R. Pettibone, O. S. Mabrouk, V. L. Hetrick, R. Schmidt, C. M. V. Weele, R. T. Kennedy, B. J. Aragona, and J. D. Berke, "Mesolimbic dopamine signals the value of work," 2015. [Online]. Available: https://www.nature.com/articles/nn.4173

[102] T. Sugiyama, N. Schweighofer, and J. Izawa, "Reinforcement learning establishes a minimal metacognitive process to monitor and control motor learning performance," 2023. [Online]. Available: https://www.nature.com/articles/s41467-023-39536-9

[103] F. Nardi, S. Haar, and A. A. Faisal, "Bill-evr: An embodied virtual reality framework for reward-and-error-based motor rehab-learning," 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10304742

[104] R. S. Sutton and A. G. Barto, "Reinforcement learning: An introduction," 2018. [Online]. Available: http://www.incompleteideas.net/book/the-book-2nd.html

[105] A. S. Therrien and A. L. Wong, "Mechanisms of human motor learning do not function independently," 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnhum.2021.785992/full

[106] J. Baladron, J. Vitay, T. Fietzek, and F. H. Hamker, "The contribution of the basal ganglia and cerebellum to motor learning: A neurocomputational approach," 2023. [Online]. Available: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1011024

[107] D. M. Wolpert and Z. Ghahramani, "Computational principles of movement neuroscience," 2008. [Online]. Available: https://www.nature.com/articles/nn1100_1212

[108] M. C. Tresch, P. Saltiel, and E. Bizzi, "The construction of movement by the spinal cord," 1999. [Online]. Available: https://www.nature.com/articles/nn0299_162#Sec3

[109] A. P. Georgopoulos, "Higher order motor control," 1991. [Online]. Available: https://www.annualreviews.org/doi/10.1146/annurev.ne.14.030191.002045?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub++0pubmed

[110] F. Mussa-Ivaldi, "Do neurons in the motor cortex encode movement direction? an alternative hypothesis," 1988. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/3173781/

[111] C. Ghez, J. W. Krakauer, R. L. Sainburg, and M. F. Ghilardi, *Spatial Representations and Internal Models of Limb Dynamics in Motor Learning. In: The new cognitive neurosciences.* The MIT Press, 2000.

[112] T. A. Martin, J. G. Keating, H. P. Goodkin, A. J. Bastian, and W. T. Thach, "Throwing while looking through prisms. i. focal olivocerebellar lesions impair adaptation," 1996. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/8813282/

[113] M. A. Smith and R. Shadmehr, "Intact ability to learn internal models of arm dynamics in huntington's disease but not cerebellar degeneration," 2005. [Online]. Available: https://journals.physiology.org/doi/full/10.1152/jn.00943.2004

[114] C. M. Harris and D. M. Wolpert, "Signal-dependent noise determines motor planning," 1998. [Online]. Available: https://www.nature.com/articles/29528

[115] M. A. S. Laith Alhussein, "Motor planning under uncertainty," 2021. [Online]. Available: https://elifesciences.org/articles/67019

[116] D. Kleinman, S. Baron, and W. Levison, "A control theoretic approach to manned-vehicle systems analysis," 1971. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1099842?casa_token=8hXOCg2UeLgAAAAA:zHB9DxBvvnithreVh4l58q5Z0BlYg0SHWvQeNw8qRVdlITtcd1WLumEKD1GYgHkbRQ2lWeeF

[117] S. Cheng and P. N. Sabes, "Modeling sensorimotor learning with linear dynamical systems," 2006. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2536592/

[118] R. J. Baddeley, H. A. Ingram, and R. C. Miall, "System identification applied to a visuomotor task: Near-optimal human performance in a noisy changing task," 2003. [Online]. Available: https://www.jneurosci.org/content/23/7/3066

[119] J. F. Soechting and M. Flanders, "Errors in pointing are due to approximations in sensorimotor transformations," 1989. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/2769350/

[120] E. Todorov, "Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system," 2004. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1550971/

[121] F. J. Valero-Cuevas, M. Venkadesan, and E. Todorov, "Structured variability of muscle activations supports the minimal intervention principle of motor control," 2009. [Online]. Available: https://journals.physiology.org/doi/full/10.1152/jn.90324.2008

[122] J. Buzzi, E. D. Momi, and I. Nisky, "An uncontrolled manifold analysis of arm joint variability in virtual planar position and orientation telemanipulation," 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8370066

[123] P. Vassiliadis, G. Derosiere, C. Dubuc, A. Lete, F. Crevecoeur, F. C. Hummel, , and J. Duque, "Reward boosts reinforcement-based motor learning," 2021. [Online]. Available: https://www.cell.com/iscience/fulltext/S2589-0042(21)00789-6?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS2589004221007896%3Fshowall%3Dtrue

[124] R. J. van Beers, E. Brenner, and J. B. J. Smeets, "Random walk of motor planning in task-irrelevant dimensions," 2013. [Online]. Available: https://journals.physiology.org/doi/full/10.1152/jn.00706.2012?rfr_dat=cr_pub++0pubmed&url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org

[125] M. K. Holden, "Virtual environments for motor rehabilitation: review," 2005. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/15971970/

[126] F. A. dos Santos Mendes, J. E. Pompeu, A. M. Lobo, K. G. da Silva, T. de Paula Oliveira, A. P. Zomignani, and M. E. P. Piemonte, "Motor learning, retention and transfer after virtual-reality-based training in parkinson's disease-effect of motor and cognitive demands of games: a longitudinal, controlled clinical study," 2012. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/22898578/

[127] T. Eggert, D. Y. P. Henriques, B. M. 't Hart, and A. Straube, "Modeling inter-trial variability of pointing movements during visuomotor adaptation," 2021. [Online]. Available: https://link.springer.com/article/10.1007/s00422-021-00858-w

[128] E. Schulz and S. J. Gershman, "The algorithmic architecture of exploration in the human brain," 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959438818300904?via%3Dihub

[129] "Unity real time development platform," https://unity.com/.

Appendix A

Appendix

A.1. Reaching Task Intertrial Variability Analysis

Analysis of the reward modulation of the intertrial variability for the reaching task with reward feedback.

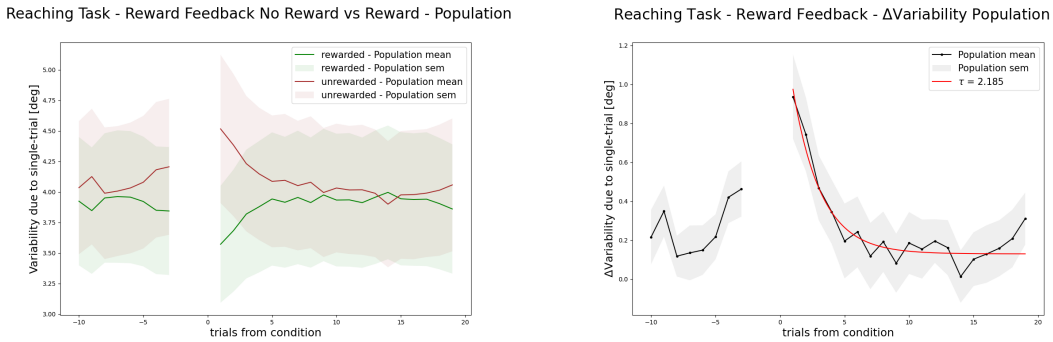


Figure A.1.: Variability following a conditioned trial average on the reaching task population

Analysis of the success modulation of the intertrial variability for the reaching task with error feedback.

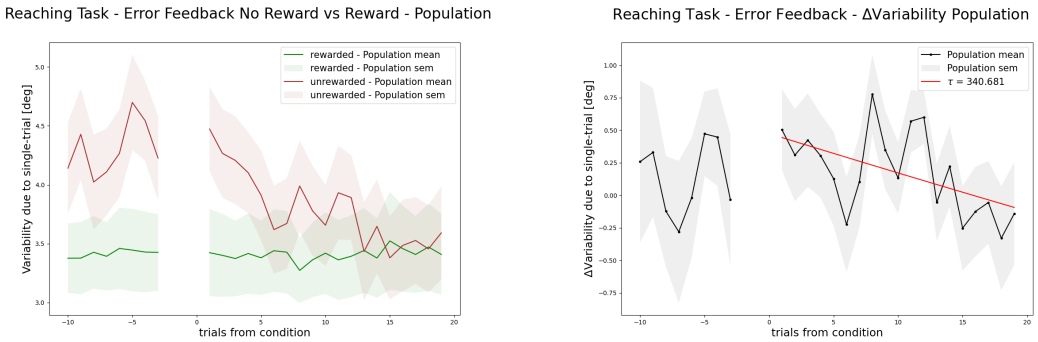


Figure A.2.: Variability following a conditioned trial average on the reaching task population

A.2. Behaviour Block Analysis

A.2.1. Pool Task

Note: The following plots exploring the behaviour at a block level, including the variability, the success rate and the lag-1 autocorrelation on the pool task have been made by Federico Nardi. This analysis is relevant in the context of this project, and allows us to interpret the results of the model.

Upon the introduction of the perturbation, a sudden increase in movement variability and a decrease in the success rate is typically exhibited. As participants gradually adapt to the perturbation, we expect a reduction in variability and a subsequent increase in the success rate.

Furthermore, corrections to the current movement are made based on an error observed at the past trial, which implies that there will be a relation between the angles of consecutive movements.
To this end, the lag-1 Autocorrelation ACF(1) has been introduced as a skill learning metric [91]. This metric quantifies the correlation between consecutive trials. ACF(1) is positive when the two angles are correlated, whereas an ACF(1) of zero implies that the two consecutive points are independent. Thus, theoretically, we start the adaptation phase with a positive ACF(1) that decreases towards zero as the perturbation is learnt and no correction is needed anymore.
The following block analysis of the Variability, Success rate and lag-1 autocorrelation was conducted by Federico Nardi on the pool task dataset. To observe the overall trends while mitigating the effect of noise, the metrics were first aggregated across blocks of 25

trials for each participant and then averaged across all participants.
The variability is calculated as the corrected standard deviation of each block (i.e. corrected with respect to the linear regression on the block).

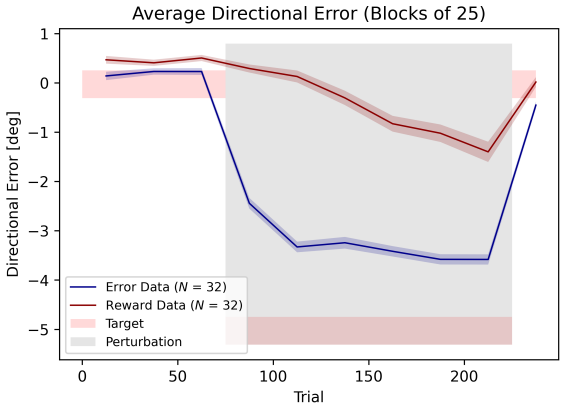


Figure A.3.: Cue Ball Angles aggregated over blocks of 25 trials and averaged across participants in each condition

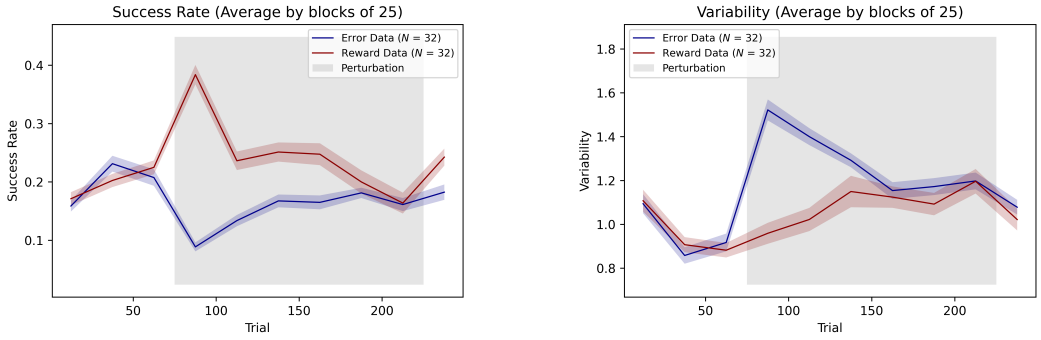


Figure A.4.: Success Rate (left) and variability (right) per block over the population in the Pool task

In the context of error feedback, the observed behaviour aligns with our expectations. Following the introduction of the perturbation, the success rate (i.e. actual pocketed ball, unbeknownst to the participant) decreases, while variability increases. As the adaptation phase progresses, the success rate gradually improves, and variability decreases as the participant learns to accommodate the perturbation.

Conversely, in the case of reward feedback, the observed behaviour may not align with

our expectations at first. This is explained by the specific way in which the task is defined.
The dynamic reward zone in the pool task was defined to gradually lead participants towards the actual reward zone.
The upper bound of this dynamic reward zone is defined as the median of the past ten rewarded trials until the actual reward zone is reached.
However, this introduces a bias that we need to account for. Early on, obtaining rewards is relatively straightforward since the reward zone is initially quite extensive. This is notably easier compared to the baseline, where rewards were granted for pocketing the target ball (i.e. approximately $[-0.311, 0.254]$ funnel around the optimal angle). However, as the participant learns to accommodate to the perturbation, the upper bound of the reward zone decreases and thus obtaining a reward becomes harder(see **??**). As the reward zone decreases, the participant needs to keep exploring and thus the variability does not decrease.

In a traditional context, the decay in autocorrelation from the early phase to the final phase of adaptation serves as a measure of learning. In the pool task, this decline is observed by computing the difference between the ACF(1) in the first half and the ACF(1) in the second half of the perturbation phase. The distribution of this decay across participants for both the error and the reward condition is shown below.

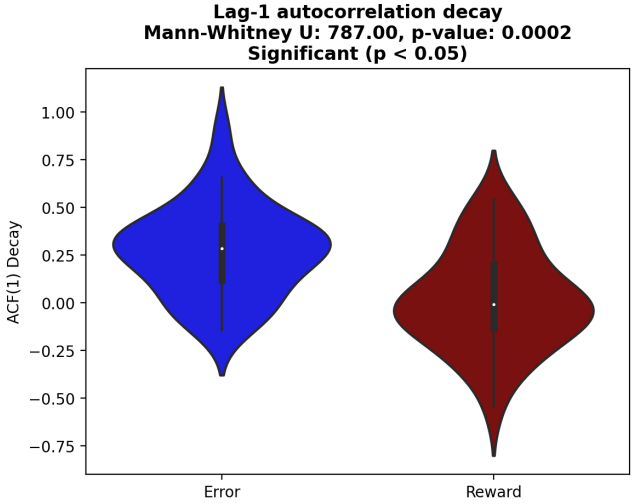


Figure A.5.: Decay in ACF(1) averaged over blocks observed over the population in the Pool task

The decay in lag-1 autocorrelation is in line with our previous observations. In the error condition, we observe clear evidence of learning.
In the reward condition, due to the specific definition of the reward zone, there is no decrease in lag-1 autocorrelation. Note that this does not imply that learning is absent

(see **??**).

A.2.2. Reaching task

A comparable analysis was conducted on the reaching task to serve as a baseline comparison with a more conventional task. It's noteworthy to mention a variation in the definition of the dynamic reward zone2.2, specifically using the median of the past ten angles rather than the median of the past ten rewarded angles. Consequently, the dynamic reward zone exhibited an occasional upward trend instead of a constant decrease for some subjects. To address this, we manually discarded subjects who may have been influenced by this distinct definition.

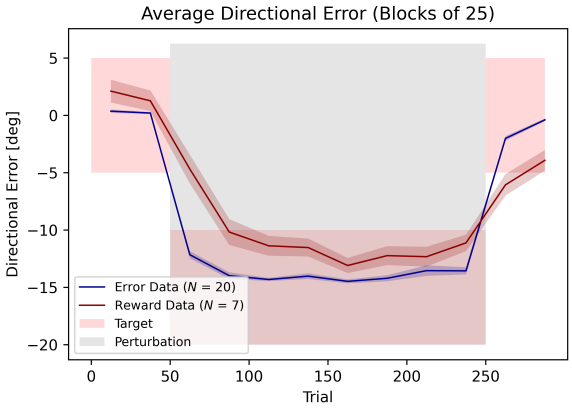


Figure A.6.: Reach angles aggregated over blocks of 25 trials and averaged across participants in each condition

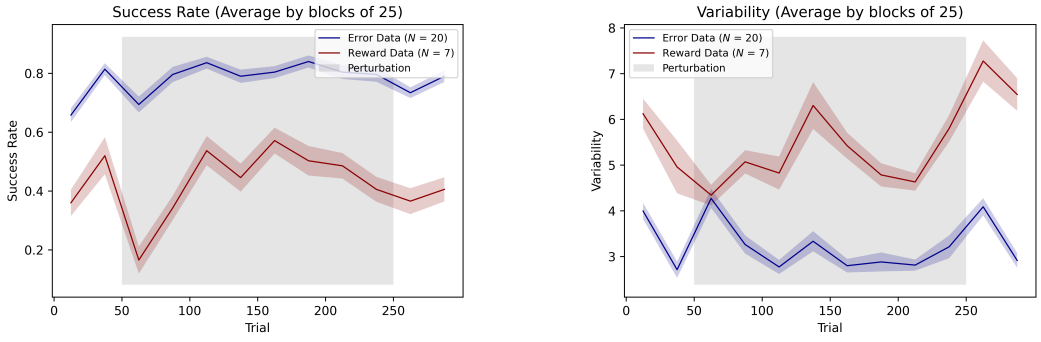


Figure A.7.: Success Rate (left) and variability (right) per block over the population in the Reaching task

In the reaching task, the steady-state reward zone is considerably larger than in the pool task (±5). As a result, in the reward condition, the reward rate is higher in the steady state, and the inherent variability of natural movements falls within this range (see A.4.4). As a result, the introduction of the dynamic reward zone does not enhance the reward rate. However, as the perturbation is introduced and the zone gradually shifts, it tends to decrease the reward rate, prompting the participant to explore again until the adapted reward zone is reached. Subsequently, variability decreases again.

Notably, the variability in the reaching task is overall higher compared to the pool task. This disparity can likely be attributed to the perturbation magnitude, with the reaching task featuring a more substantial perturbation (15° as opposed to 5°) and subsequently more important changes in the motor command.
Furthermore, we notice an increase in variability in both feedback conditions around the middle of the perturbation phase. This peak is likely due to the set breaks forgetting effect [98] and is more pronounced in the steady state phase.

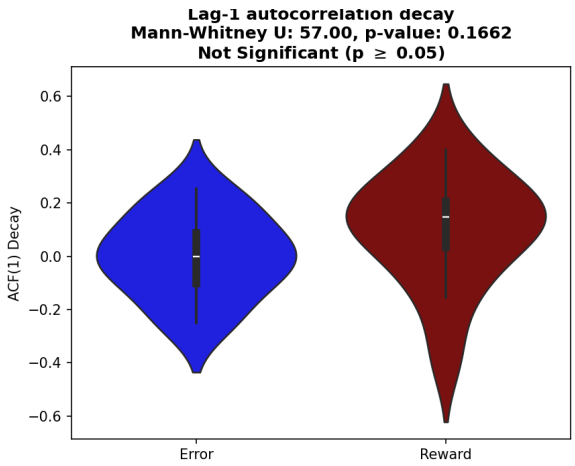


Figure A.8.: Decay in ACF(1) averaged over blocks over the population in the reaching task

We witness a nearly negligible decay in the error-feedback case. This can be intuitively explained by the fact that the reaching task is significantly less complex(i.e. hand reaching movement versus pocketing a ball), and is reflected in both the high success rate A.11 and the participant examples A.4.4. In the error condition, participants adapt almost instantaneously to the perturbation (i.e. $1 - 10$ trials).
In the reward condition, consistent with the variability observations, we see a decay in the lag-1 autocorrelation, meaning that the participants reduce their variability between the first and the second half of the perturbation phase.

A.3. Model Results - Reaching Task

A.3.1. Error-based Model

Once we have fitted the parameters to each participant, we run one model simulation per subject and compare the average resulting behaviour between the experiments and our model's prediction.

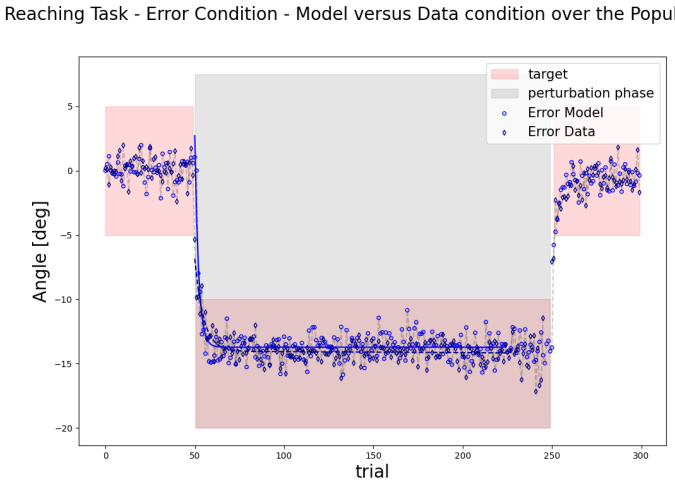


Figure A.9.: Reaching task - Model fitted to Subject vs Subject's Data averaged over the population

The distribution of the parameters is the following:

Error Model Parameters - Reaching task								
Parameters	a_p		σ_p^2		σ_h^2	σ_m^2		
Value	0.623	$\pm$	1.844	$\pm$	1.84 ± 1.438	1.809	$\pm$	
	0.482		1.438			1.415		

We compare the average directional error, success rate and variability aggregated over blocks of 25 trials, averaged across participants:

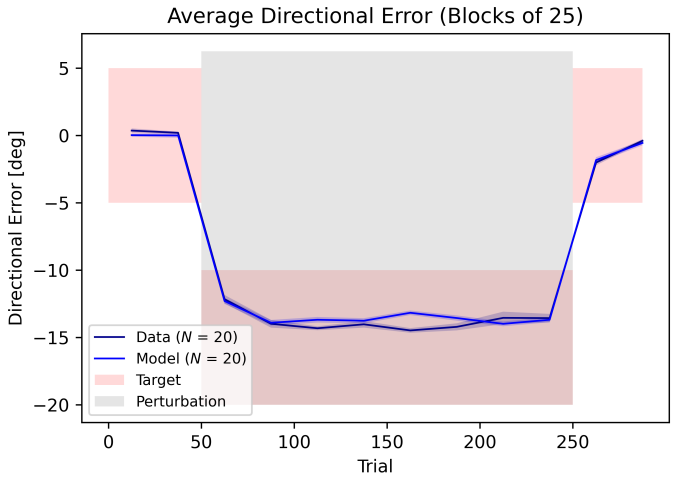


Figure A.10.: Reaching task - Model fitted to Subject vs Subject's Data over the population

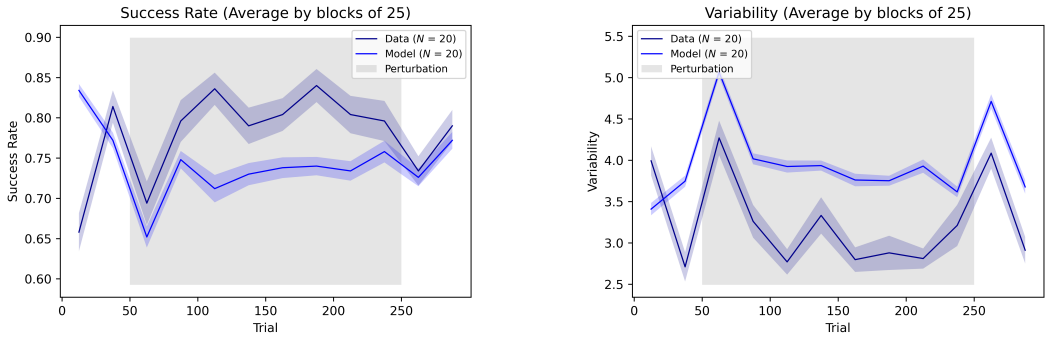


Figure A.11.: Success Rate (left) and variability (right) per block over the population in the Reaching task

We observe the following performance metrics after fitting the model to each of the participants.

Error Model Performance - Reaching task			
Metrics	$RMSE$	R^2	$WMSE$
Value over Population	1.516	0.947	0.102
Value over Subject	4.358 ± 3.649	0.292 ± 0.247	0.262 ± 0.221

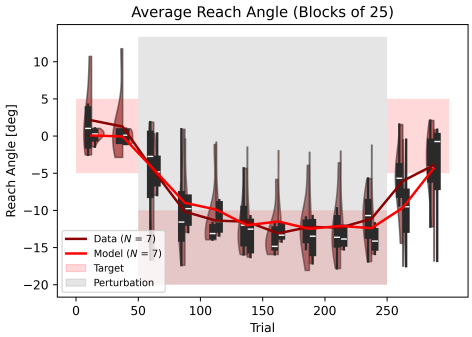


Figure A.12.: Comparison of the block averaged angle of the Reward Model and the participants on the reaching task in the reward condition.

Note that it is important to be cautious when making comparisons between the model's performances on the pool and reaching task due to the significant disparity in perturbation magnitude, leading to considerably different levels of observed movement variability.

A.3.2. Reward-based Model

By employing the fitting method introduced earlier 6.2, we note that the reward model effectively encompasses a broad range of behaviours displayed by participants in the reaching task. In this task, the overall high levels of variability and relative readiness to obtain success mitigate the between-subject distinction of variability and adaptation behaviour. The substantial overall noise level makes achieving a precise fit challenging, but we can discern that the model captures various trends.
We observe distinct adaptation rates, varying levels of overall variability, as well as different modulations of variability following reward conditions. Additionally, the oscillation patterns resulting from the set breaks are effectively captured by the reward model.

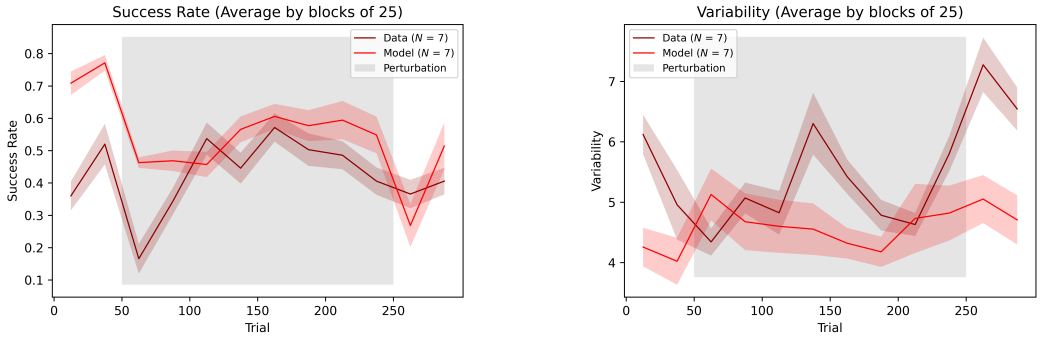


Figure A.13.: Success Rate (left) and variability (right) per block over the population in the Pool task

We note the following distribution of parameters:

Reward Model Parameters - Reaching task						
Parameters	α_x	α_r	σ_m^2	σ_p^2	λ	a
Value	0.255 ± 0.087	0.812 ± 0.127	0.923 ± 0.158	2.128 ± 0.536	2.375 ± 0.414	6.75 ± 1.601

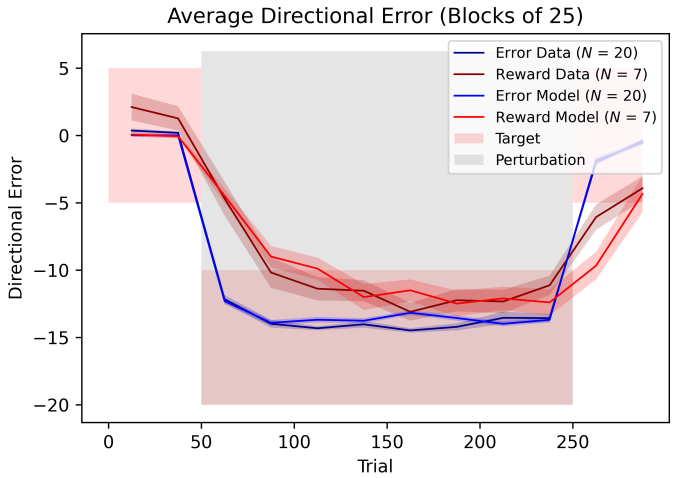


Figure A.14.: Comparison of the average angle between the Error and Reward models and the participant's data in the reaching task

We observe the following performances after fitting the model to each of the selected participants.

Reward Model Performance - Reaching task			
Metrics	$RMSE$	R^2	$WMSE$
Value over Population	3.983	0.602	0.237
Value over Subject	7.337 ± 0.785	0.345 ± 0.032	0.484 ± 0.089

We notice a relatively low RMSE, considering the extent of variability present in the reward condition of the reaching task. Note that this value is lower when comparing the average behaviours of the model and the participants over the population, as opposed to calculating the average RMSE between each participant and their corresponding fitted model. This is explained by the larger number of repetitions, which results in the emergence of a more resilient overall pattern that is less susceptible to noise.
The reasonably high R-squared (R2) value indicates that the model effectively accounts for a substantial portion of the variability present in the data.

A.4. Participants examples Model vs Data

A.4.1. Pool Task - Reward condition

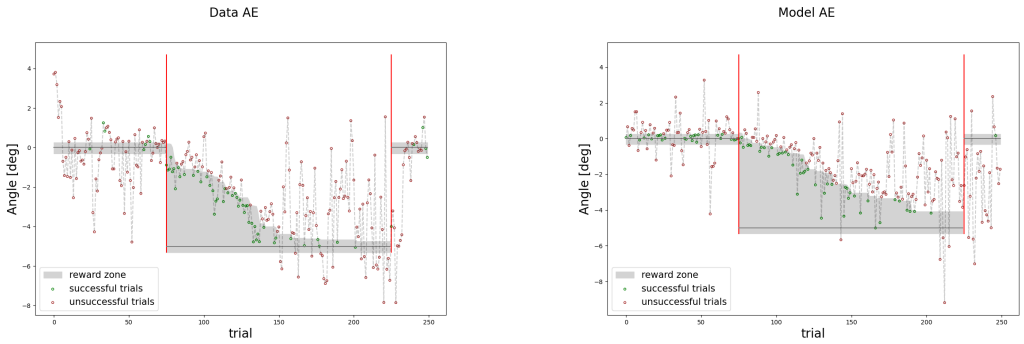


Figure A.15.: Pool task example Subject Data (left) and Reward Model simulation(right), in the reward feedback condition

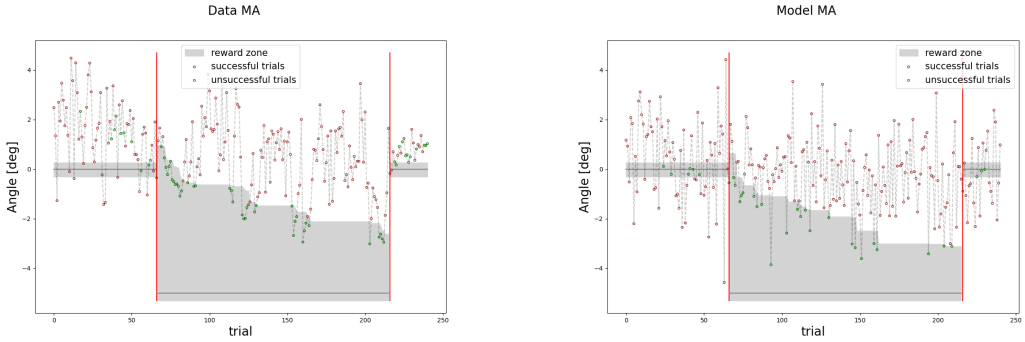


Figure A.16.: Pool task example Subject Data (left) and Reward Model simulation(right), in the reward feedback condition

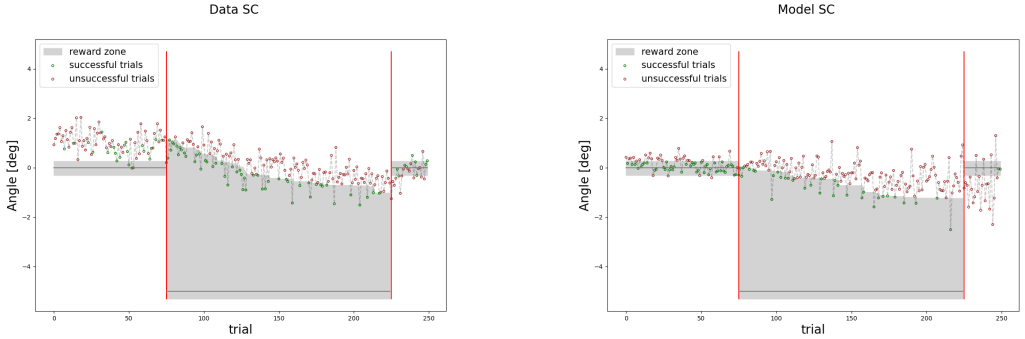


Figure A.17.: Pool task example Subject Data (left) and Reward Model simulation(right), in the reward feedback condition

A.4.2. Pool Task - Error condition

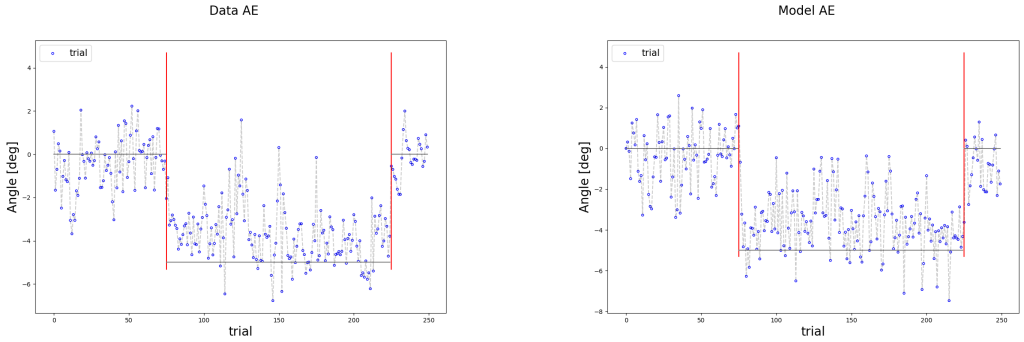


Figure A.18.: Pool task example Subject Data (left) and Error Model simulation(right), in the error feedback condition

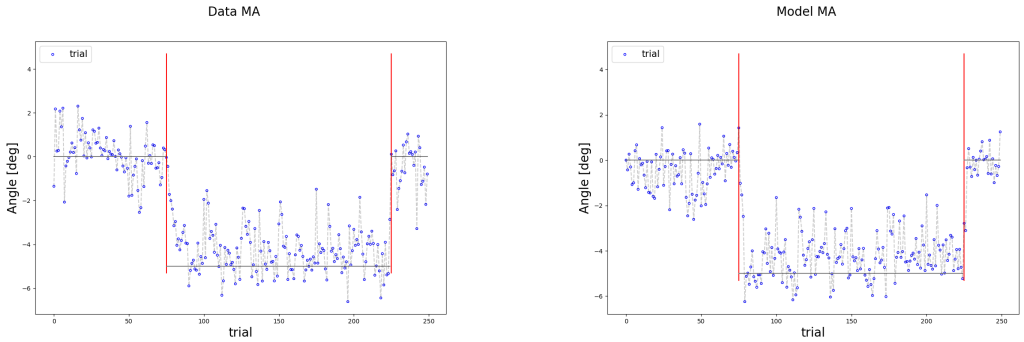


Figure A.19.: Pool task example Subject Data (left) and Error Model simulation(right), in the error feedback condition

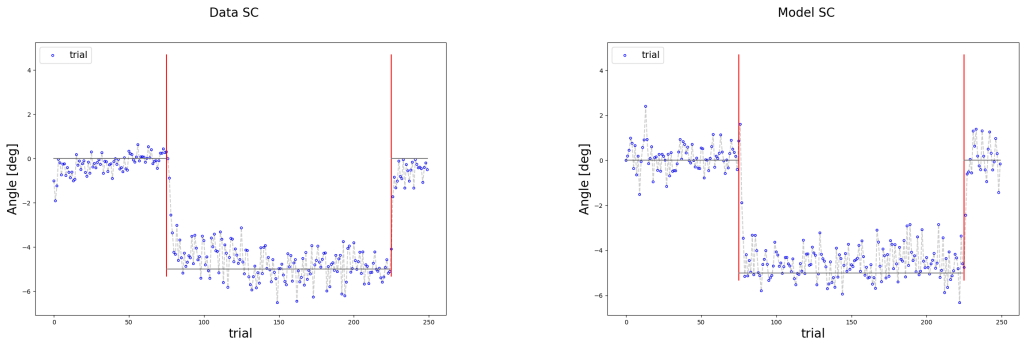


Figure A.20.: Pool task example Subject Data (left) and Error Model simulation(right), in the error feedback condition

A.4.3. Reaching Task - Reward condition

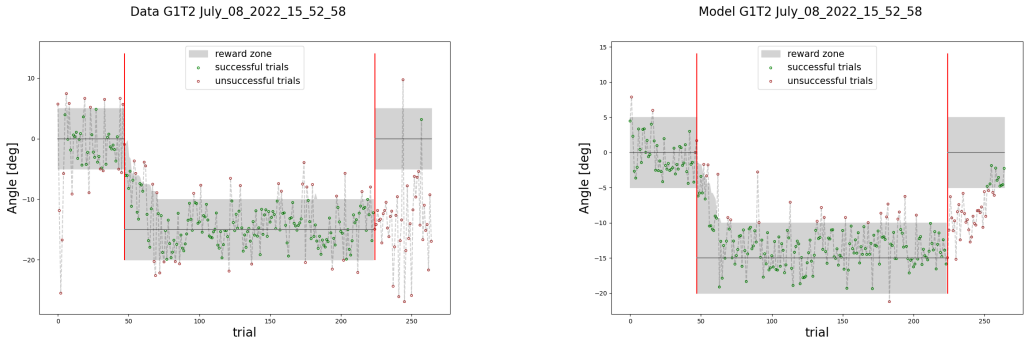


Figure A.21.: Reaching task example Subject Data (left) and Reward Model simulation(right), in the reward feedback condition

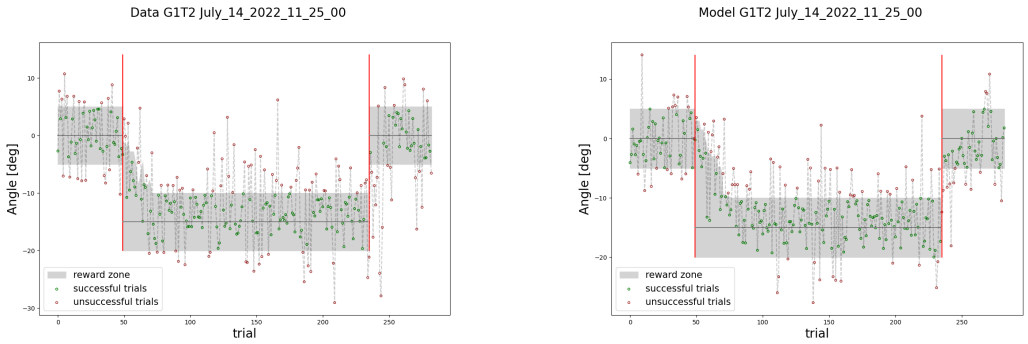


Figure A.22.: Reaching task example Subject Data (left) and Reward Model simulation(right), in the reward feedback condition

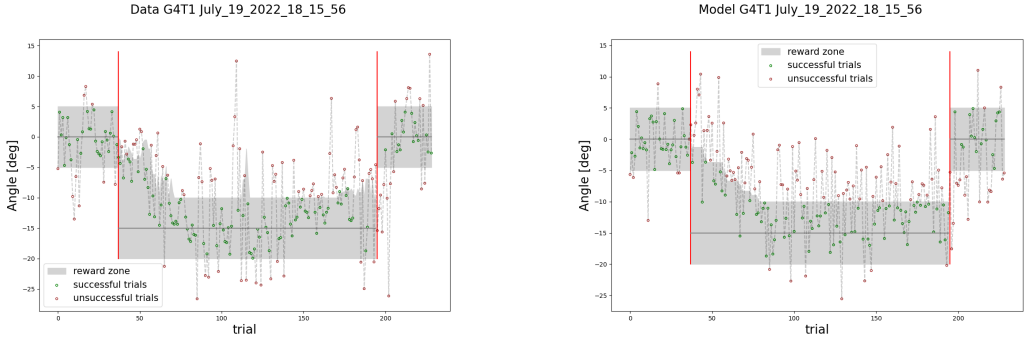


Figure A.23.: Reaching task example Subject Data (left) and Reward Model simulation(right), in the reward feedback condition

A.4.4. Reaching Task - Error condition

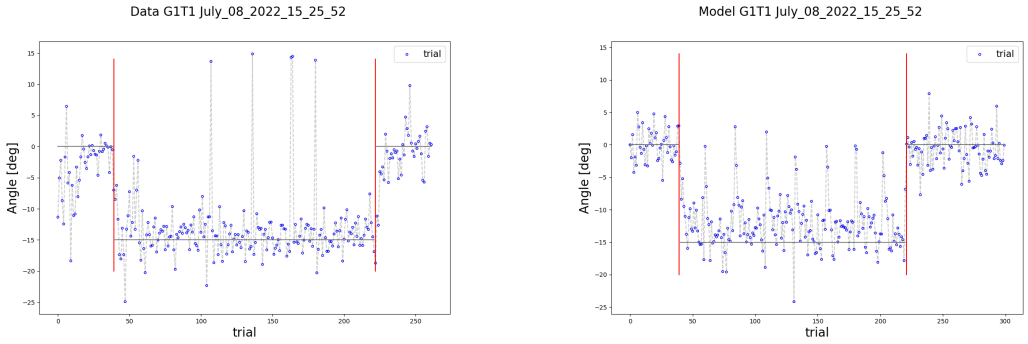


Figure A.24.: Reaching task example Subject Data (left) and Reward Model simulation(right), in the reward feedback condition

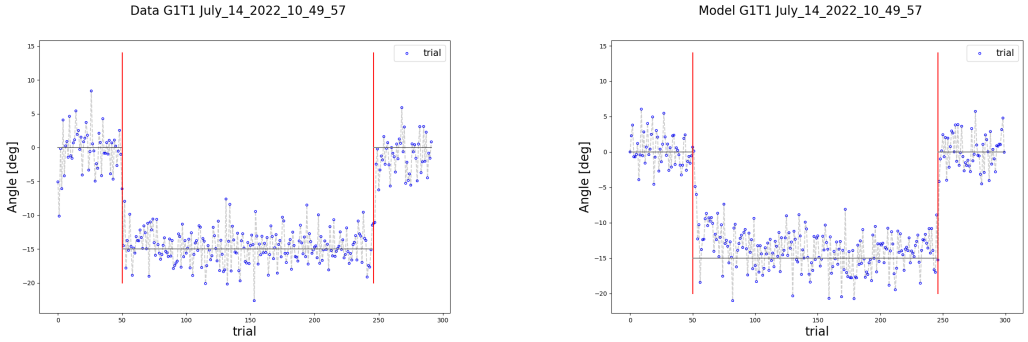


Figure A.25.: Reaching task example Subject Data (left) and Reward Model simulation(right), in the reward feedback condition

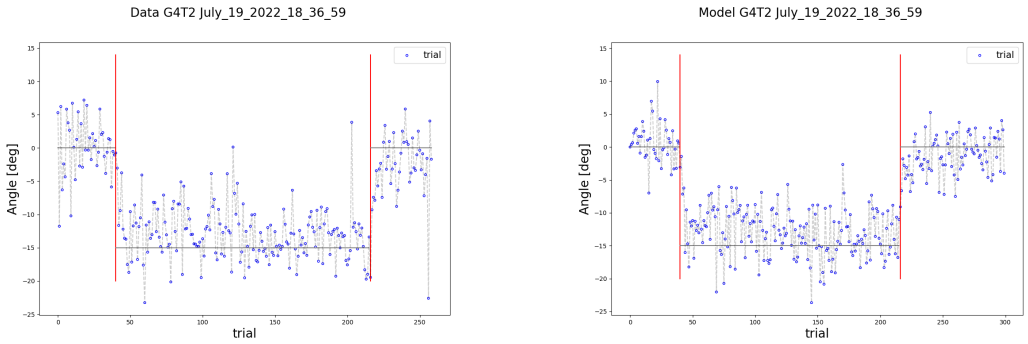


Figure A.26.: Reaching task example Subject Data (left) and Reward Model simulation(right), in the reward feedback condition

A.5. Model implementation

A.5.1. Error Model

```
    class KalmanFilter():
    def __init__(self, transition_matrix, input_matrix,
transition_covariance,
                 observation_matrix, observation_covariance,
                 initial_state_mean, initial_state_covariance):
        self.A = transition_matrix
        self.B = input_matrix
        self.W = transition_covariance
        self.C = observation_matrix
        self.V = observation_covariance
        self.prior = gaussian(initial_state_mean,
initial_state_covariance)

    def log_likelihood(self, z):
        P = np.dot(np.dot(self.A,self.prior.cov),np.transpose(self.A)) +
self.W
        S = np.dot(np.dot(self.C,P),np.transpose(self.C)) + self.V
        return logpdf(z, np.dot(self.C, self.prior.mean), S)

    def get_estimate(self, a, y_observed, A, B):
        predicted_estimate = np.dot(A ,self.prior.mean.reshape(-1,1)) +
np.dot(B,a)
        predicted_covariance = np.dot(np.dot(A,self.prior.cov),np.
transpose(A)) + self.W
        # error (residual) between measurement and prediction
        innovation_estimate = y_observed - np.dot(self.C,
predicted_estimate)
        # project system uncertainty into measurement space
        innovation_covariance =  np.dot(np.dot(self.C,
predicted_covariance),np.transpose(self.C)) + self.V
```

```
24
25          # Kalman gain is the weight given to the innovation (ie., the
    difference between the measurement and the predicted measurement)
26          # Weight different sources of uncertainty
27          try:
28              K = np.dot(np.dot(predicted_covariance ,np.transpose(self.C))
    ,np.linalg.inv(innovation_covariance))
29              I = np.eye(K.shape[0])
30          except:
31              K = np.dot(predicted_covariance ,np.transpose(self.C)) /
    innovation_covariance
32              I=1
33          # predict new x with residual scaled by the kalman gain
34          updated_mean = predicted_estimate + np.dot(K,innovation_estimate)
35
36          updated_cov = np.dot((I - np.dot(K,self.C)),predicted_covariance)
37          posterior = gaussian(updated_mean, updated_cov)
38
39          # Current posterior becomes next-step prior
40          self.prior = posterior
41          return posterior.mean.reshape(-1,)
```

Listing A.1: Kalman Filter Estimator

```
1     class LQG():
2   def __init__(self, transition_matrix, input_matrix,
    transition_covariance,
3                observation_matrix, observation_covariance,
    initial_state_mean,
4                initial_state_covariance, state_cost, control_cost,
    state_dim=2, ntrials=1, static_noise=False, T=300):
5     self.A = transition_matrix
6     self.B = input_matrix
7     self.W = transition_covariance
8     self.C = observation_matrix
9     self.V = observation_covariance
10    self.Q = state_cost
11    self.R = control_cost
12    self.static_noise = static_noise
13    self.ntrials = ntrials
14    self.t = 0
15    self.latent_states = np.zeros((state_dim,T, ntrials))
16    self.latent_states[:,0] = (initial_state_mean + np.random.
    multivariate_normal([0.0,0.0], initial_state_covariance)).reshape
    (-1,1)  #np.random.normal(0.0, self.W)
17    self.T = T
18
19  def step(self, action, A, B, t):
20    self.t=t
21    measurement = action + np.random.normal(0.0, self.W[1,1]) + np.random
    .normal(0.0, self.V) + self.perturbation[t]
22    return measurement
23
24
```

```
25  def get_control_gain(self, A,B):
26
27    P_LQR = solve_discrete_are(A, B, self.Q, self.R)
28    #Intermediate variable
29    S = np.dot(np.dot(np.transpose(B),P_LQR),B)
30    S2 = np.dot(np.dot(np.transpose(B),P_LQR),A)
31    # Control feedback gain
32    L = np.dot(np.linalg.inv(self.R + S),S2)
33
34    return L
35
36  def control_policy_LQG(self, control_gain, esimated_state):
37
38    current_action =  -np.dot(control_gain,esimated_state)
39    return current_action
```

Listing A.2: LQG Actor

```
1   def simulate_error_model(A_p, A_h, W_p, W_h, V, Q, d, init_state,
      block_trial_index, perturbation_phase_index, washout_phase_index,T
      =300,  x_dim=2, visuomotor_perturbation=15, savedir=None, plotting=
      False, compute_likelihood=False, angle_data=None):
2     A_init = np.array([[A_p,0],[A_h,0]])
3     B_init = np.array([0,1]).reshape(-1,1)
4     C = np.array([0,1]).reshape(1,-1)
5     W = np.array([[W_p,0],[0,W_h]])
6     V = np.array([V])
7     R = np.array([1-Q])
8     Q = np.array([[0,0],[0,Q]])
9     ini_state_mean = init_state
10    ini_state_cov = np.array([[.1,0],[.1,0]])
11    estimates = np.zeros((x_dim,T))
12    estimates[:,0] = ini_state_mean
13    x_dim=x_dim
14    #Time between break sets
15    end_block_index = block_trial_index
16    d = d
17    perturbation = np.zeros(T)
18    perturbation[perturbation_phase_index:washout_phase_index] =
      visuomotor_rotation
19
20    estimator = KalmanFilter(transition_matrix=A_init,
21                         input_matrix=B_init,
22                         transition_covariance=W,
23                         observation_matrix=C,
24                         observation_covariance=V,
25                         initial_state_mean=ini_state_mean,
26                         initial_state_covariance=ini_state_cov)
27
28    actor = LQG(transition_matrix=A_init,
29                    input_matrix=B_init,
30                    transition_covariance=W,
31                    observation_matrix=C,
32                    observation_covariance=V,
```

```
                    initial_state_mean=ini_state_mean,
                    initial_state_covariance=ini_state_cov,
                    state_cost = Q,
                    perturbation=perturbation,
                    control_cost = R,
                    state_dim=x_dim,
                    static_noise=False,
                    T=T)

    action = np.zeros(T)
    y_feedback = np.zeros(T)
    model_output = np.zeros(T)
    measurement = np.zeros(T)
    log_likelihood = np.zeros(T)
    x = np.zeros((x_dim,T))
    for t in range(1, T):
        if t in block_trial_index:
            A = np.power(A_init,1/d+1)
            B = B_init
        else:
            A = A_init
            B = B_init
        control_gain = actor.get_control_gain(A,B)
        measurement[t] = actor.step(action[t-1],A,B, t)
        #Introduce perturbation

        y_feedback[t] = measurement[t]
        estimates[:,t] = estimator.get_estimate(action[t-1],y_feedback[t
    ],A,B)
        if compute_likelihood:
            log_likelihood[t] = estimator.log_likelihood(angle_data[t])

        action[t] =  actor.control_policy_LQG(control_gain, estimates[:,t
    ])

    model_output = action
    if compute_likelihood:
        avg_log_likelihood = sum(log_likelihood)
        return model_output, avg_log_likelihood
    else:
        return model_output
```

Listing A.3: Error Model

A.5.2. Reward Model

```
    class Critic():
    #Approximate value function on continuous action space by calculating
     exponentially weighted reward history
    def __init__(self, tau = 1.2):
        super(Critic, self).__init__()
        # reward rate
```

```
6         self.alpha_r = 1-np.exp(-1/tau)
7         self.avg_rwd_rate = 0.5
8         self.avg_rwd_rate_s = 0.5
9     def reward_prediction_error(self, reward):
10
11        delta_t = reward  - self.avg_rwd_rate
12        self.avg_rwd_rate = self.avg_rwd_rate + self.alpha_r * delta_t
13        #No negative reward rate
14        if self.avg_rwd_rate < 0:
15             self.avg_rwd_rate = 0
16        return delta_t
```

Listing A.4: Critic

```
1   class Actor():
2     def __init__(self, mu_init = 1, exploration_noise_init = 1,
      execution_noise = 2, planning_noise = 0.2, variability_max = 6,
      variability_decay = 0.3, alpha_mu = 0.3, d = 1):
3         super(Actor, self).__init__()
4
5         # mean estimation
6         self.alpha_mu = alpha_mu
7         self.mu = mu_init
8         self.d = d
9         #fixed motor execution noise
10        self.sigma_m = execution_noise
11        #fixed planning noise
12        self.sigma_p = planning_noise
13        #Regulated exploratory noise is modulated by the value function (
      reward rate estimate)
14        self.sigma_e = exploration_noise_init
15        self.variability_max = variability_max
16        self.variability_decay = variability_decay
17
18        self.mu_s = self.mu
19
20    def get_action(self, execution_noise = True):
21        #Exploration variability is defined by a fixed planning noise and
       some regulated exploration noise
22        exploration_noise = np.random.normal(0.0, self.sigma_e) + np.
      random.normal(0.0, self.sigma_p)
23        if execution_noise:
24            angle = self.mu  + exploration_noise + np.random.normal(0.0,
      self.sigma_m)  #+ self.mu_s)/2
25        else:
26            angle = self.mu + exploration_noise
27        return angle, exploration_noise
28
29    def variability_function(self,x):
30        return self.variability_max*np.exp(-x/self.variability_decay)
31
32    def update_exploratory_noise(self, avg_rwd_rate):
33        rwd_regulated_variability = self.variability_function(
      avg_rwd_rate)
```

```
        self.sigma_e = rwd_regulated_variability

        #No negative variability
        if self.sigma_e < 0:
             self.sigma_e = 0


    def update_mean(self, delta_t, n_e, set_break = False):
        if set_break:
            self.mu = self.mu*self.d**2 + 1/self.d*self.alpha_mu*delta_t*
    n_e
        else:
            self.mu = self.mu + self.alpha_mu*delta_t*n_e
```

Listing A.5: Acotr

```
class reward_model():
    def __init__ (self, mu_init = 1, exploration_noise_init = 1,
    execution_noise = 2, planning_noise = 0.2, variability_max = 6,
    variability_decay = 0.3, alpha_mu = 0.3, d = 1, tau = 1.2, num_trials
    = 300, task = 1):
        super(reward_model, self).__init__()
        # Reward calculation
        self.last10Angles = np.zeros(10)
        self.task = task
        # Store mean estimates
        self.mu_memory = np.zeros(num_trials+1)
        self.mu_memory[0] = mu_init
        self.n_e_memory = np.zeros(num_trials+1)
        self.n_e_memory[0] = exploration_noise_init

        self.actor = Actor(mu_init, exploration_noise_init,
    execution_noise, planning_noise, variability_max, variability_decay,
    alpha_mu, d)
        self.critic = Critic(tau)

        self.t = 0
    def update_median(self, angle, reward=None):
        if reward == 1:
                self.last10Angles = np.roll(self.last10Angles, shift=len(
    self.last10Angles)-1)
                self.last10Angles[-1] = angle

    def forward(self, perturbation=False, set_break=False,
    execution_noise=True):
        self.t += 1
        # Set Break reset reward rate estimate to 0.5 (neutral)
        if set_break:
            self.critic.avg_rwd_rate = 0.5
        # Actor generates angle
        angle, n_e = self.actor.get_action(execution_noise)
        # Environment gives us the reward
        reward, dyn_reward_zone = self.compute_reward(angle, task = self.
    task, perturbation_bool = perturbation)
```

```
        self.update_median(angle, reward)
        # Critic computes the Value function (reward rate estimate)
        delta_t = self.critic.reward_prediction_error(reward)
        # Update the regulated planning Variability for exploration
        self.actor.update_exploratory_noise(self.critic.avg_rwd_rate)
        self.n_e_memory[self.t] = self.actor.sigma_e
        #update the mean estimate od the angle
        self.actor.update_mean(delta_t, n_e, set_break)
        self.mu_memory[self.t] = self.actor.mu
        return angle, reward, dyn_reward_zone

    def compute_reward(self, angle, task, perturbation_bool=False):
        if task == 1:
            perturbation_val=5
            optimal_angle = 0
            lbPocket = optimal_angle-0.3106383
            ubPocket = optimal_angle+0.2543617
            corrPerturbation = -5
        else:
            perturbation_val=15
            optimal_angle = 0
            lbPocket = optimal_angle-5
            ubPocket = optimal_angle+5
            corrPerturbation = -15
        reward=0
        if perturbation_bool:
            M = median(self.last10Angles, 10)
            ubPocket = ubPocket + corrPerturbation
            lbPocket = lbPocket + corrPerturbation

            if angle < np.maximum(M,ubPocket) and angle > lbPocket:
                reward = 1
            dyn_reward_zone = np.maximum(M,ubPocket)
        else:
            dyn_reward_zone = ubPocket

            if angle < ubPocket and  angle > lbPocket:
                reward = 1
        return reward, dyn_reward_zone
```

Listing A.6: Reward Model