An Examination into the Putative Mechanisms Underlying Human Sensorimotor Learning and Decision Making

Jack Brookes

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

University of Leeds
School of Psychology
Leeds, UK

December 2019

Intellectual Property and Publications

The candidate confirms that the work submitted is his/her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The following publications were obtained during the course of this PhD:

Chapter 2 includes work that is under review and has been posted as a pre-print (*Exploring Disturbance as a Force for Good in Motor Learning*, Jack Brookes, Faisal Mushtaq, Earle Jamieson, Aaron J. Fath, Geoffrey P. Bingham, Peter Culmer, Richard M. Wilkie, Mark A. Mon-Williams, bioRxiv 796136; doi: https://doi.org/10.1101/796136). The methods were originally devised by, and majority of the data was originally presented by, Earle Jamieson (Jamieson, E. S., 2015. Haptic Enhancement of Sensorimotor Learning for Clinical Training Applications, PhD, University of Leeds). The work in Chapter 2 is the author's own and arose from discussions involving collaborators.

Chapter 4 includes work from a jointly authored publication (*Studying human behavior with virtual reality: The Unity Experiment Framework*, Jack Brookes, Matthew Warburton, Mshari Alghadier et al. Behav Res (2019); doi: https://doi.org/10.3758/s13428-019-01242-0). The author wrote the manuscript, developed the framework, and performed analyses for the case study. The data from the case study were collected by M. Alghadier. Collaborating authors commented on sections of the manuscript and edits in response to these comments are included in this chapter.

Acknowledgements

I would like to express my gratitude to my supervisors Faisal Mushtaq, Pete Culmer & Andrew Keeling for the assistance during my PhD. In particular, Faisal has gone above and beyond supervisor duties, taking time out of evenings and weekends to help overcome research difficulties and provided career advice. In addition, Mark Mon-Williams deserves a special mention for offering invaluable mentoring and big-picture thinking. I would not have been able to complete this body of work without the support I have received from the ICON and PACLab groups, and especially those who frequent lab room LG29. I am eternally thankful to Mshari Alghadier, Zeynep Uludag, Awais Hafeez, Matthew Warburton, Emily Stanyer, Emily Robinson, Rashaun Black & Darnell Sparkes-Wallace for their valiant data collection efforts.

As the first in my immediate family to go to University, I should thank my parents Paul & Louise for encouragement and support through this (seemingly never-ending) process. Finally, I am forever in debt to my partner Emily who has been unconditionally supportive throughout my write-up.

Overview

Sensorimotor learning can be defined as a process by which an organism benefits from its experience, such that its future behaviour is better adapted to its environment. Humans are sensorimotor learners par excellence, and neurologically intact adults possess an incredible repertoire of skilled behaviours. Nevertheless, despite the topic fascinating scientists for centuries, there remains a lack of understanding about how humans truly learn. There is a need to better understand sensorimotor learning mechanisms in order to develop treatments for individuals with movement problems, improve training regimes (e.g. surgery) and accelerate motor learning in tasks such as handwriting in children and stroke rehabilitation. This thesis set out to improve our understanding of sensorimotor learning processes and develop methodologies and tools that enable other scientists to tackle these research questions using the power of recent developments in computer science (particularly immersive technologies). *Errors* in sensorimotor learning are the specific focus of the experimental chapters of this thesis, where the goal is to address our understanding of error perception and correction in motor learning and provide a computational understanding of how we process different types of error to inform subsequent behaviour. A brief summary of the approaches employed, and tools developed over the course of this thesis are presented below.

Chapter 1 of this thesis provides a concise overview of the literature on human sensorimotor learning. It introduces the concept of internal models of human interactions with the environment, constructed and refined by the brain in the learning process. Highlighted in this chapter are potential mechanisms for promoting learning (e.g. error augmentation, motor variability) and outstanding challenges for the field (e.g. redundancy, credit assignment).

In *Chapter 2* a computational model based on information acquisition is developed. The model suggests that disruptive forces applied to human movements during training could improve learning because they allow the learner to sample more information from their environment. *Chapter 3* investigates whether sensorimotor learning can be accelerated through forcing participants to explore (and thus acquire more information) a novel workspace. The results imply that exploration may be a necessary component of learning but manipulating it in this way is not sufficient to accelerate learning. This work serves to highlight the critical role of error correction in learning.

The process of conducting the experimental work in Chapters 2 and 3 highlighted the need for an application programme interface that would allow researchers to rapidly deploy experiments that allow one to examine learning in a controlled but ecologically relevant manner. Virtual reality systems (that measure human interactions with computer generated worlds) provide a powerful tool for exploring sensorimotor learning and their use in the study of human behaviour is now more feasible due to recent technological advances. To this end, *Chapter 4* reports the development of the Unity Experiment Framework - a new tool to assist in the development of virtual reality experiments in the Unity game engine.

Chapter 5 builds on the findings from Chapters 2 & 3 on learning by addressing the specific contributions of visual error. It utilises the Unity Experiment Framework to explore whether visually increasing the error signal in a novel aiming task can accelerate motor learning. A novel aiming task is developed which requires participants to learn the mapping between rotations of the handheld virtual reality controllers and the movement of a cursor in Cartesian space. The results show that the visual disturbance does not accelerate the learning of skilled movements, implying a crucial role for mechanical forces, or physical error correction, which is consistent with the findings reported in Chapter 2. Uncontrolled manifold analysis provides insight into how the variability in selected solutions related to learning and performance, as the task deliberately allowed a variety of solutions from a redundant parameter space.

Chapter 6 extends the scope of this thesis by examining how error information from the sensorimotor system influences higher order action selection processes. Chapter 5 highlighted the loose definition of “error” in sensorimotor learning and here, the goal was to advance our understanding of error learning by discriminating between different sources of error to better understand their contributions to future behaviour. This issue is illustrated through the example of a tennis player who, on a given point, has the options of selecting a backhand or forehand shot available to her. If the shot is ineffective (and produces an error signal), to optimise future behaviour, the brain needs to rapidly determine whether the error was due to poor shot selection, or whether the correct shot was selected but just poorly executed.

To examine these questions, a novel ‘action bandit’ task was developed where participants made reaching movements towards targets, with each target having distinct probabilities of execution and selection error. The results revealed a significant selection bias towards a target that produced a higher frequency of execution errors (rather than a target associated with more selection error) despite no difference in expected value. This behaviour may be explained by a gating mechanism, where learning from the lack of reward is discounted following sensorimotor errors. However, execution errors also increase uncertainty about the appropriateness of a selected choice and the need to reduce uncertainty could equally account for these results. Subsequent experiments test these competing hypotheses and show this putative gating mechanism can be dynamically regulated though coupling of selections and execution errors. Development of models of these processes highlighted the dynamics of the mechanisms that drive the behaviour. In Chapter 7, the motor component of the task was removed to examine whether this effect is not unique to execution errors, but a feature of any two-stage decision-making process with, multiple error types which are presumed to be dissociated. These observations highlight the complex role error plays in learning and suggest the credit assignment process is guided and modulated by internal models of the task at hand.

Finally, Chapter 8 closes this thesis with a summary of the key findings and arising from this work in the context of the literature on motor learning and decision making.

It is noted here that this thesis sought to cover two broad research topics of motor learning and decision making that have, until recently, been studied by separate groups of researchers, with very little overlap in literature. A key goal of this programme of research was to contribute towards bringing together these hitherto disparate fields by focussing on breadth to establish common ground. As the experimental work developed, it became clear that the processing of error required a multi-pronged approach. Within each experimental chapter, the focus on error was accordingly narrowed and definitions refined. This culminated in developing and testing how individuals discriminate between errors in the sensorimotor and cognitive domains, thus presenting a framework for understanding how motor learning and decision making interact.

Notes

Unless otherwise noted, error bars represent +/-1 standard error of the mean.

Data collection for this programme of work was approved by the Psychology Research Ethics Committee at the University of Leeds (Approved: 25/10/16, Ethical Approval Number: 16 – 0269).

Contents

List of Figures

1. Introduction

The first chapter in this thesis is designed to provide an overview of key topics, concepts and theories that will be covered in the methods and experimental chapters in the remainder of this thesis. The purpose is to set the scene and provide readers with an introduction to core concepts to be examined in the chapters that follow, rather than an exhaustive summary of the intricacies of motor learning (which would be beyond the scope of a single thesis).

1.1. Principles of Sensorimotor Control and Learning

Movement is a fundamental feature of life. Animals have a remarkable ability to perform a range of complex movements with incredible precision. Motor control is not a unitary process, it spans the integration of both external (e.g. visual, auditory) and internal (proprioception) sensory information (Todorov, 2004), planning a desired movement trajectory (Bizzi et al., 1984; Harris and Wolpert, 1998), recruitment of necessary muscles and execution of the planned trajectory using muscle contractions (Bernshteĭn, 1967).

1.1.1. Perception

Perception is the process that allows the motor system to obtain the necessary information required to generate successful execution plans (Grush, 2004). It involves, for example, filtering and transforming raw signals (such as photons hitting the retina, or vibrations in the ear) into information that can be classified and usefully interpreted by the brain to make sense of the world around us (Hommel, 2005; Shadmehr et al., 2010).

Importantly, extraction of sensory information is not a passive process. Instead, movements can be used to *change* the information we gather, in order to extract the most important information from our environments (Brown et al., 2013). For example, visual information acquired can be altered by orienting the eyes towards targets of interest, as required by the planning process of the motor system (Wolpert et al., 2011). These types of behaviours are highly stereotyped in neurologically intact individuals and performed in a near optimal Bayesian fashion for information extraction (Najemnik and Geisler, 2005).

1.1.2. Internal models

Perception is used within a feedback loop in order to generate and adjust execution plans. In order effectively generate these action plans required to meet our goals, humans likely hold internal models of their bodies and the external world (Francis and Wonham, 1976; Kawato, 1999; Kawato and Wolpert, 2007). A forward model is used to generate predictions of the consequences of events in the world. The most obvious example is the consequence of our own actions – for example, we use a forward model to predict the sensory consequences of a reaching movement. It might take the current angles and velocities of the arm joints and output an estimate of the subsequent future arm joint angles and velocities. Taken one step further, a separate forward model might take the predicted arm joint angles and velocities and produce a prediction of the sensory information that this state would produce. Forward models also are used to make predictions about the outside world. For example, an approaching ball's position and velocity can be used in tandem with features of the world such as acceleration due to gravity in order to make predictions about the ball's future position (Wolpert and Miall, 1996).

In contrast, an inverse model generates a required set of events that would lead to a given state (Wolpert and Kawato, 1998). For motor control, this could be the joint angles that are required in order to position the hand at a desired location. Again, these models could be modular and hierarchical, such that the required joint angles in this example must then also pass through an inverse model, along with other state variables such as current angles and velocities, to produce the muscle contractions necessary to produce the required joint angles (Diedrichsen and Kornysheva, 2015).

Commonly an inverse model would be comprised of a many-to-one mapping, such that any given state could be caused by any number of different event sequences (Wolpert and Miall, 1996). This creates the so-called "Degrees of freedom" problem in which the motor system must select an execution strategy from a large search space (Bernshteĭn, 1967), discussed later. These internal models are fundamentally malleable, as our body composition changes through growth or injury, or the environment changes (e.g. rain may cause a slippery path), the internal models can be refined to accommodate the changes (Wolpert, 1997).

1.1.3. Implementation

In order to execute a complex action, the brain may employ a combination of a set of fundamental "motor primitives" (Mussa-Ivaldi et al., 1994; Thoroughman and Shadmehr, 2000; Flash and Hochner, 2005; Diedrichsen and Kornysheva, 2015; Giszter, 2015). These primitives represent the muscle activity pattern of the most basic movements performed by the body. The motor hierarchy begins with a high-level selection of the action required to achieve a task. Then, the required motor primitives are selected and activated in order to produce the instructed command. There is also evidence for "chunking", that is, intermediate groupings of motor primitives that are activated in tandem, which can simplify the selection of motor primitives and speed up reaction time (Lashley, 1951, pp.112–146; Diedrichsen and Kornysheva, 2015).

Wolpert (2000; 2010; 2011) poses three classes of motor control. Predictive/feedforward control employs forward models to make predictions about the future state of the world, and then uses inverse models to generate actions which achieve the desired goal. The potential delays that can be experienced in the sensorimotor system make predictions key to performing actions (Miall et al., 1993; Franklin and Wolpert, 2011). Reactive control modifies currently executing actions in response to new feedback which was not predicted by the feedforward control mechanism. This is clearly essential as the predictions are not entirely accurate and so are subject to error, and adjustments may be needed to perform the desired action (Wolpert and Kawato, 1998). Finally, biomechanical control modulates stiffness of a limb in order to mitigate interference from external perturbations or even to dampen internal noise (Wolpert et al., 2011).

1.1.4. Learning from action

The ability to improve one’s motor skills through learning is a crucially important skill for humans and other animals alike. Motor learning has been defined as any experience-dependent improvement in performance (Krakauer et al., 2019) and is a blanket term that encompasses many observable phenomena.

A widely cited proposal by Fitts and Posner (1967) on human skill acquisition separates the process of learning a motor task into three distinct stages: the cognitive, associative, and autonomous stages. To illustrate, consider the processes involved in learning to ride a bicycle. The cognitive stage might entail learning the high-level, explicit rules that govern the system, pushing down on the pedals with one’s legs harder when climbing a hill, and squeezing the brakes with one’s hand to slow down when approaching a turn. During the associate stage, some of these rules become memorised, and no longer does the learner have to think which leg to push with in order to speed up. Here, the learner might begin to think about the minutia of the human–bicycle system, how hard the brakes should be pulled, the weight distributions when turning and so on. Finally, in the autonomous stage, these processes become automatic and no longer have to be given high-level thought by the rider, and only small refinements are made (Taylor and Ivry, 2012).

1.1.5. Error based learning

When an action is executed, e.g. a reaching to pick up a cup of coffee, we can measure the resulting position or trajectory of our hand. We can use this measurement and compare it to the expected position or trajectory of our hand and form an error vector (position and direction). Then, subsequent movements can be adjusted in an attempt to correct for this error.

Errors are thought to be corrected through a process of gradient descent, where the system attempts to walk the parameter space (representing constituent control parameters, e.g. joint angles) “downhill”, in order to minimise this error (Sailer et al., 2005; Mosier et al., 2005; Johansson et al., 2006; Wolpert et al., 2011). However, the error correction rate (or step size) must be tuned such that learning is sufficiently fast, but not overly fast that the system attempts to correct for errors that are the result of inherent noise in perception, planning, or execution (van Beers, 2009).

Error-based learning is responsible for the effects seen in motor adaptation experiments, where errors are induced in, for examples, reaches through the introduction of perturbations. Perhaps the most widely used and oldest approach are “visuomotor transformations” – where visual information is manipulated, e.g. offset by a constant amount to induce an error signal (Helmholtz, 1867; Welch, 1978; Shadmehr and Mussa-Ivaldi, 1994; Krakauer et al., 2000; Morehead et al., 2015) and must be corrected by the participant to successfully perform the task. Physical analogues of this approach have become increasingly more common with the introduction of haptics, where directional forces must be counteracted though muscle contractions for the user to achieve their goal (Shadmehr and Mussa-Ivaldi, 1994; Lee and Choi, 2010; Heuer and Lüttgen, 2015).

1.1.6. Reinforcement learning

Error signals typically provide information on the direction and magnitude of the discrepancy between current and desired state, which the motor system can use to update subsequent motor plans. However, error-based learning is relevant only in task-space. That is, it helps to correct for errors made by the end-effector (e.g. the hand) but cannot change how the various constituent control parameters (e.g. joint angles) are used in combination across a redundant space. Changes across this redundant space (or “solution manifold”) will not decrease error on average, as all combinations are valid solutions for meeting the target, but some solutions may induce more noise than others.

To change the solution used across this redundant parameter space, we must employ a form of reinforcement learning (Barto et al., 1998). Reinforcement learning can facilitate exploration of this space – if errors are still present after the error-based learning mechanism, reinforcement learning will facilitate de-selection of that movement “strategy”, in favour of others. Reinforcement learning is especially useful where sequences of actions are carried out, and the presence or lack of reward assigns credit or blame to the preceding actions (Wolpert et al., 2011).

1.1.7. Model based / model free

Motor learning is generally thought to comprise of two distinct processes (Sutton and Barto, 1998; Haith and Krakauer, 2013), (1) a model-based process which improves performance through development of internal models using prediction errors (Wolpert and Miall, 1996), and (2) a simpler model-free process which occurs within the controller, reinforcing the use of certain actions through reward feedback (Rescorla and Wagner, 1972). Both are thought to contribute towards the learning of any skilled task, with both processes developing in parallel (Daw et al., 2011). Model-based learning may be more important in the early stages, where there is no useful internal model that can be deployed for the task at hand. When performance has developed sufficiently through development of internal models, model-free learning can refine the model. This reflects the fact that habitual learning often develops later in the learning process (Balleine and O'Doherty, 2010).

1.1.8. Structural learning

A closely related concept to model-based learning is a process known as structural learning. The structure of a task (e.g. riding a bicycle) can be thought of as the mathematical relationships between the relevant inputs (e.g. fundamental human movements) and outputs (e.g. movements of the bicycle) (Wolpert et al., 2011). In the case of riding a bike, it is clear that the structure is dependent on some internal and external factors, such as the mass of our arms and the width of the handlebars. If this internal structure is known and understood, one can easily generalise any skill learned to a similar task (e.g.

riding a differently shaped bike). If this structure is not known, performance can still be good on the bike that was practised on, but it will be difficult to generalise to a variant of this task. Studies have shown that structural learning can be facilitated by allowing participants to practice a task with varying parameter values between trials. Then, when an assessment is performed, even with parameter values that have been previously unseen, participants see an improved ability to generalise their learning to the new task (Braun et al., 2009).

1.1.9. Variability, noise and learning

Even the most skilled movements are subject to variability, as there is noise present in each step of the system, from the uncertainty in location of a target through our senses, to the noise in executing movements. This inherent variability has often been seen as an undesirable side-effect of movement. Indeed, many societies prize individuals who show exceptional ability to minimise motor variability – from golfers and darts players to football. Yet, variability may also act as a facilitator of learning (Dhawale et al., 2017) and indeed a number of recent studies have emerged proposing that this motor noise can be beneficial to the learning process (Tumer and Brainard, 2007; Wu et al., 2014).

Variability may be usefully separated into task-relevant and task-irrelevant variability (Wolpert et al., 2011), but the relationship each of these has separately on motor learning is not well understood. A recent analysis on variability in learning found mixed results on how variability impacts motor learning (He et al., 2016), indicating a need for more empirical evidence on the subject.

1.1.10. Redundancy in movement

The majority of movements made by humans involve only a subset of an infinite possible selection of movements that could be made to achieve the same goal. This point was first described by Bernshteĭn (1967) and referred to as the degrees of freedom problem. Bernshteĭn proposed that the many joints in the human body have the ability to be grouped into a "synergy", where many multiple degrees of freedom into a fewer number (Li, 2006). Alternatively, degrees of freedom may be eliminated completely by locking down joints/freezing, e.g. through rigid fixation of multiple joints (Vereijken et al., 1992; Berthouze and Lungarella, 2004), which is often observed in the early stages of learning (Newell, 1991).

In general, the degrees of freedom problem can be thought of as a parameter search in a high dimensional space where many solutions exist that meet a set of constraints. These techniques that reduce the effective degrees of freedom essentially place more constraints on the solution, perhaps making the solution easier to acquire (Li, 2006).

1.1.11. The Uncontrolled Manifold Hypothesis

Related to the degrees of freedom problem, the Uncontrolled Manifold (UCM) hypothesis (Scholz and Schöner, 1999; Latash et al., 2001; Latash et al., 2010; Scholz and Schöner, 2014) defines a subspace (manifold) of the parameter space (i.e. all possible joint angles) which have different joint angles yet still meet the demands of the task (e.g. desired hand position). Motion within this manifold would not affect the control variables (those that affect the given task) and is therefore unnecessary to control, thus the name “uncontrolled”. This concept helps provide a language to describe the complex joint space, by defining the uncontrolled manifold that *does not* affect the task, and an orthogonal subspace that *does* affect the task.

Synergies, redundant groups of joints executed in tandem to produce a motor trajectory, can be identified through examining the position of solutions along this manifold (Latash et al., 2001). The UCM concept allows for calculation of two different types of variability – task-space (variability that affects outcomes) and null-space (variability that does not affect outcomes). It is not known how these two types of variability inter-relate throughout the learning process. The goal of motor learning is reduction of task-space variability (Wolpert et al., 2011) but could facilitate or reduce null-space variability without any impact on performance (Cardis et al., 2017).

The causal influences of these possible mechanisms were recently investigated by Cardis et al. (2017), who found that any addition of artificial variability (task-space or null-space) hindered performance. However, correlative analysis may still reveal how humans’ natural variability could help in the exploration of new solutions. Indeed, Singh et al. (2016) through UCM analysis found a significant relationship between observed null-space variability and subsequent learning rates.

1.1.12. Exploration vs exploitation

A related concept is the exploration vs exploitation trade-off, studied first in animal and human learning and now an important consideration in machine learning (Kaelbling et al., 1996). Any learning system is constantly faced with the dilemma of whether to explore new possible strategies, or exploit existing known strategies to maximise long-term reward. Intuitively, an optimal strategy would consist of early stages being dominated by exploration, and once sufficient information has been accumulated about the environment, later stages taking an exploitation strategy that attempts to maximise utility from interacting with the environment. Additionally, the reward level offered should play a role – one might expect more exploratory behaviour when playing tennis casually versus a friend, but refrain from exploring when partaking in a competition with a large prize (Dhawale et al., 2017).

Several models have been proposed for understanding how humans resolve this dilemma, with work being done to understand the neural mechanisms of these processes (Daw et al., 2006; Boorman et al., 2009; Raja Beharelle et al., 2015). Distinct behaviours have been observed in various tasks with non-stationary reward schedules. Participants seem to be sensitive to changes in rewards but seem to either attempt to switch strategies in hope of finding an action that grants a greater reward, or double-down and try harder with the current strategy, depending on the task (Rabbitt, 1966; Laming, 1979; Gratton et al., 1992; Cohen et al., 2007). Additionally, humans are sensitive to the predicted length of time of the task (Carstensen et al., 1999). Generally, belief that the task will take a long time leads to early exploratory behaviour, since there is more time to reap the rewards of the exploration (Cohen et al., 2007).

1.1.13. Cognitive & motor interactions

Although often viewed as separate systems, recent studies have explored how higher level cognitive processes (e.g. economic decision making) and lower level motor control processes (action planning, execution) interact (McDougle, Boggess, et al., 2016; Parvin et al., 2018). In the real world, all decisions are implemented through execution of actions to some extent, and thus the relatively little crosstalk between the motor learning and decision-making worlds has been rather surprising –a state of affairs that has only recently started to be addressed (Taylor et al., 2014; Chen et al., 2018; Aczel et al., 2018; Codol et al., 2018).

The general principle that higher level action selection, or cognitive, strategies cannot be considered without taking into account the processes involved in implementing those actions (i.e. sensorimotor control) have long been promoted by embodied theorists who propose a bilateral relationship between action and cognition (Wilson, 2002). Some elegant examples of these cognitive-motor interactions can be found in the visuomotor rotation literature (McDougle, Ivry, et al., 2016; Holland et al., 2018; Codol et al., 2018). Motor adaptation to changes in the environment were previously assumed to exclusively involve implicit process of refining a forward model using a prediction error signal. Humans are however able to utilise explicit processes such as verbal instructions to assist in their learning. Taylor et al. (2014) separated the implicit and explicit processes of adaptation by looking at the difference between verbal reports of aiming direction as well as actual measured aiming direction. They concluded that sensorimotor adaptation is a result of the interplay between implicit and explicit processes, after observing both explicit learning, achieved by initial high exploratory behaviour, and slower and more consistent implicit adaptation.

Explicit reward systems can also impact motor learning. Indeed, Galea at al. (2015) showed that there is an asymmetry between the effects or reward and punishment on motor learning, with punishment on errors facilitating accelerated learning, and rewards on successful movements facilitating memory retention. Chen et al. (2018) reiterate these findings, and highlight a need for novel experiments where action selection and action execution are dissociated, and reward/punishment mechanisms are examined in relation to both stages.

1.2. Challenges to address

The preceding sections have provided introductions to core concepts and now we discuss some of the challenges in the field of human sensorimotor learning that will be tackled as we navigate through this thesis. In particular, this thesis aims to further understand the mechanisms by which humans are able to refine their action selection and execution abilities through learning from *errors*. Errors can provide not only a magnitude of reward or punishment that indicates to the user how well they performed the movement, but also a direction which can be used to refine future actions. Each experimental chapter aims to build on this research by testing hypotheses about the mechanisms of error-based learning.

1.2.1. Assistive, disruptive forces & positional control

Force intervention can be used to alter the motor learning process, but we have limited underlying mechanism driving this phenomenon (Sigrist et al., 2013). Positional control is a force intervention technique with full movement assistance delivered via some kind of robot (requiring no muscle control from the user). The assistive device moves the limb along a pre-defined trajectory in order to teach the user an optimal movement (Feygin et al., 2002; Sigrist et al., 2013). This intervention technique prevents the user from making their own errors, which are crucial to motor learning (Emken and Reinkensmeyer, 2005). However, this technique may be useful where the user has impaired movement abilities, or is in the very early stages of learning, just as a parent might guide a child's hand to teach them to write their first words (Sigrist et al., 2013).

Haptic guidance (or assistance) is similar to positional control, but with levels of force intervention that do still require some level of effort from the user to complete the task. The intervention applies a force in the direction of a path or target such that less effort is required by the user than if there was no intervention. These types of intervention have been successfully applied to those with neurological conditions impairing movement (Marchal-Crespo and Reinkensmeyer, 2009) but can hamper the motor learning of skilled subjects (Cesqui et al., 2008; Sigrist et al., 2013).

Disruptive forces applied during movements have been shown to accelerate the learning of a motor task (Emken and Reinkensmeyer, 2005; Reinkensmeyer and Patton, 2009; Milot et al., 2010; Williams et al., 2016). Explanations of the processes that drive this phenomenon have largely focussed on the idea of increased attentional allocation driven by error increased error(Marchal-Crespo et al., 2017) or alternatively, impedance control – suggesting individuals learn a strategy of stiffening their arm to mitigate external disturbances, which reduces error even when these disturbances are subsequently removed (Takahashi et al., 2001; Sigrist et al., 2013). Alternative untested hypotheses are that the disruptive forces facilitate exploration of the task space, therefore promoting model-based learning of the task structure, or that error-based learning is enhanced through a larger (on average) positional error.

Together with haptic guidance, it is clear that there is no one-size-fits-all approach to enhancing motor learning using force interventions, but a promising approach may be to modulate the level of assistance/disruption based on skill level, in order to deliver an optimal balance between error enhancement and instructional mechanisms (Rauter et al., 2010; Sigrist et al., 2013; Kahn et al., 2014). Additionally, application of these techniques to areas such as sport, surgery or dental training is a promising avenue to help shorten the long learning process that some professions require (Reinkensmeyer and Patton, 2009). The mechanism behind these phenomena are investigated in Chapters 2 & 3.

1.2.2. Error amplification through visual manipulation

As illustrated by visuomotor transformation experiments (Helmholtz, 1867; Jeannerod et al., 1995; Flanagan and Rao, 1995; Kitazawa et al., 1997), error can be manipulated visually in addition to external force perturbations. This manipulation can help dissociate the roles of the *perception* of error and the *correction* of error on learning. Recent experiments have shown participants are able to improve performance beyond a previously defined ceiling level when exposed to a visual error amplification intervention in a virtual throwing task (Hasson et al., 2016) and reaching (Patton et al., 2013). However, a recent application of this intervention to a rowing task revealed no benefit (Gerig et al., 2019). Noteworthy is the fact that none of these studies focus on the performance after-effects of these interventions, i.e. is the performance improvement retained once the intervention is removed? This is important because this intervention is one that can potentially applied to rehabilitation (Wei et al., 2005), sport (Milanese et al., 2008), and surgical training systems (Reinkensmeyer and Patton, 2009) providing a lasting benefit that can be used in the real world. Additionally, experiments on visual error amplification can help understand the mechanistic link between errors and motor learning that have been observed though enhancement of errors through disruptive forces. We examine the possibility that these types of interventions can accelerate learning in Chapter 5.

1.2.3. The credit assignment problem

A topic that exists in the broader field of learning concerns how humans solve various forms of the credit assignment problem (Smith et al., 2006; Kording et al., 2007; Huang and Shadmehr, 2009; Wolpert et al., 2011). The credit assignment problem concerns determining how the success of a system's overall performance is due to the numerous contributions of the system's component parts (Minsky, 1961). Humans must solve this problem many times during the learning process, when rewards are separated from the action that caused them by an amount of time, or other actions in-between. Here, it is ambiguous which action was responsible for the reward. For example, a game of chess involves many actions in a sequence; in the presence of a win or a loss, the brain must decipher which action or actions most contributed to that outcome. Proposed mechanisms on how humans solve these problems are eligibility traces (Pan et al., 2005) – which involves storing a trail of actions in memory which led to the outcome, and temporal difference learning (Stolyarova, 2018) which assigns intermediate "value" to actions in the sequence (for example, a high value might be assigned to capturing an opponent's queen piece, as it often leads to a victory). The credit assignment problem is an important concept in furthering our understanding of sensorimotor learning for two reasons. First, since motor control often involves contractions of dozens of muscles simultaneously, the brain must use some approach to credit to refine the movements from the appropriate joints. Second, the brain needs to solve the issue that errors can arise not only poor action execution, but also an incorrect action selection and this has substantial consequences on if the motor system should refine its internal models to optimise future behaviour.

It also seems intuitive that humans solve the credit assignment problem using the nature of the feedback that occurs during a reward (or lack thereof). For example, when attempting to access the reward of a caffeine hit contained in a cup of coffee, there are several scenarios which would lead to a lack of reward (no caffeine). Spilling the coffee after a poorly executed reach would lack the reward, but equally a kitchen mix-up where decaffeinated coffee was accidentally chosen instead of regular coffee would also lack the reward. However, the presence of the feedback of spilled coffee in the former scenario would help resolve ambiguity, solve the credit assignment problem, and appropriately assign blame to the poorly executed reach rather than the choice of coffee.

This type of scenario has been recently explored by making a distinction between action execution and action selection by McDougle et al. (2016). These tasks utilise multi-arm “bandit” paradigms, where a selection between several options leads to a chance of reward, with each bandit affording a chance of yielding a reward with a probability initially unknown to the participant. In a task where simple button presses are used to indicate selection of a bandit, participant elicit risk averse behaviour predicted by prospect theory (Kahneman and Tversky, 1979). However, when the required button press for selection is replaced with a reaching movement, and errors are presented as errors in execution (misses), behaviour is flipped, with a risk-seeking strategy seemingly adopted. McDougle et al. (2016) reasoned that this behaviour is the product of “gating” of higher order reinforcement learning processes, where the presence of an execution error attenuates (or gates) value updating associated with the target. Consistent with this, McDougle et al. (2019) subsequently showed that reward prediction error coding in the striatum, a subcortical region implicated in reinforcement learning, is attenuated following execution versus selection errors. In a related study, Parvin et al. (2018) showed this pattern is not driven by the strength of the sensorimotor error, but instead by the participant’s agency, or belief that they are in control of the outcomes.

Other explanations aside from this gating hypotheses, consistent with these experiments, have only minimally been explored. A key feature of an execution error in these tasks is that it results in information uncertainty concerning the potential reward if the participant had correctly executed the action. Previous explanations assume selection behaviour is driven by a desire to maximise immediate reward. However, a less certain value estimate would be held by the participant if a target were to facilitate more execution errors, as the participant has had fewer opportunities to experience the offered level of reward. Thus, this apparent risk-seeking behaviour may, at least in-part be driven by a desire to reduce this uncertainty (Cohen et al., 2007; Mushtaq et al., 2011). These ideas are explored in detail in Chapters 6 & 7.

1.2.4. Software for better science

The majority of modern scientific investigations rely on computer software to capture data and the study of human learning and decision-making is no different. After decades of speculation (Loomis et al., 1999), the potential for virtual reality technology to transform the ways in which computers are used to investigate human behaviour is now beginning to be realised. Virtual environments allow the production of novel and ecologically relevant experiments with accessible price points.

These new technologies are generally difficult to interact with, and often require detailed technical knowledge to maximise their utility. A challenge to address here is how new software can be developed to alleviate some of the technical burden placed on researchers, allowing scientific methods to be more effective, accessible, and reproducible. Chapter 4 of this thesis will introduce a new tool that scientists and educators can take advantage of in order to more readily address the types of research questions being investigated in this thesis. This tool is particularly suited for sensorimotor experiments where manipulation of feedback (e.g. error) is important, since virtual reality allows deep control of visual feedback.

2. Exploring Disruption as a Force for Good in Motor Learning

2.1. Overview

Disruptive forces facilitate motor learning, but theoretical explanations for this counterintuitive phenomenon are lacking. Smooth arm movements require predictions (inference) about the force-field associated with a workspace and these predictions require *information*. We used these insights to create a new information theory inspired model that explains why disturbance helps learning. We performed secondary analysis of data on two motor learning experiments in which participants undertook a continuous tracking task where they learned how to move their arm in different directions through a novel 3D force field. We compared baseline performance before and after exposure to the novel field to quantify learning. In Experiment 1, the exposure phases (but not the baseline measures) were delivered under three different conditions: (i) robot haptic assistance; (ii) no guidance; (iii) robot haptic disturbance. Replicating previous work, the disturbance group showed the best learning. Secondly, the nature or intensity of the error augmenting force was manipulated trial-by-trial, in an attempt to provide a skill-matched level of assistance or disruption in Experiment 2. Counterintuitively, providing an unpredictable level of assistance/disruption facilitated the most performance improvement over the skill-matched intervention. The information model was constructed in an attempt to explain these observations. By computing the amount of information acquired during learning across all experiments, 12% of the variance in learning could be explained. This account presents a new perspective on reconciling previous findings on error amplification and indicates that information may be the central currency of motor learning.

2.2. Introduction

Neonates must determine the complex relationship between perceptual outcomes and motor signals in order to learn how to move their arms effectively. This process is repeated throughout life as humans calibrate to new environments, acquire new skills, experience neuromuscular fatigue or recover from injury. Technological advances have created robotic systems designed to accelerate the acquisition of skilled arm movements in a variety of areas including, amongst others, laparoscopic surgical training and stroke rehabilitation (Reinkensmeyer and Patton, 2009). These devices can provide assistive forces that guide an individual's arm through a desired trajectory or apply disturbance forces that make it more difficult for the individual to move their arm along a given trajectory.

It is now well established that providing assistive forces to neurologically intact individuals can actually impair subsequent learning (Sigrist et al., 2013; Laura Marchal-Crespo et al., 2014). Conversely, there is growing empirical evidence that providing disruptive forces to impair performance during training of a motor task can have a net positive effect, and lead to improved learning - enhancing performance in the task after the disruptive forces are removed (Emken and Reinkensmeyer, 2005; Cesqui et al., 2008; Lee and Choi, 2010; Sigrist et al., 2013; Laura Marchal-Crespo et al., 2014; L. Marchal-Crespo et al., 2014).

However, formalised theoretical explanations that can account for these counterintuitive phenomena have proven elusive (Heuer and Lüttgen, 2015). This is disappointing because it remains unclear how robotic devices might be best optimised in order to enhance learning (beyond this binary observation of differences between assisting and disturbing forces). The lack of a theoretical framework also makes it difficult to explain formally why assistive forces can be beneficial for individuals with neurological impairment (Hesse et al., 2003), and the absence of a framework is hindering the potential utility of robotic technology in motor training. We propose that a 'Shannon'-style information theory perspective (Shannon, 1948) could provide a principled approach to understanding why disruptive forces can be beneficial, and such an account could ultimately inform the development of haptic interventions.

The development of 'forward models' that act as neural simulators regarding how the current state of the system will respond to a given motor signal (Wolpert and Miall, 1996) naturally would require repeated observation of inputs and outputs of a system. Viewed in this way, motor learning requires the system to sample information in order to extract the invariant rules that govern a range of input–output mappings (Braun, Mehring, et al., 2010; Braun, Waldert, et al., 2010). The difficulty faced by the system relates to the large number of internal parameters that connect the sensory input to the motor output i.e. high levels of uncertainty (Bays and Wolpert, 2007). The example of a neonate learning the mapping between perceptual and motor output illustrates how this problem can be framed from an information theory perspective. The new-born must use information generated from their exchanges with the environment in order to learn the input–output mappings and subsequently refine their predictions, so that they can successfully interact with their new surroundings. The initial reaches will be associated with high levels of uncertainty and thus the feedback information is of a greater value, compared to later in the learning process whereby the feedback is predictable. The developmental trajectory, however, will be marked by a reduction in entropy as the certainty of a predictable perceptual outcome following the generation of a motor command will increase. Thus, motor learning can be viewed as a process where uncertainty is reduced through the development of forward models following exposure to information regarding the relationship between perceptual output and motor signal input (Friston et al., 2010).

We propose that this information perspective can account for the previous finding of superior learning outcomes from disturbance haptic force application relative to assistive guidance. Specifically, we suggest that providing assistive forces limits information exposure and thus constrains the amount of learning. Conversely, disturbance forces expose the individual to more information which facilitates the learning process. Following this logic, a control algorithm that provides a greater level of information should lead to better learning than those that minimise uncertainty. It will be noted that a certain level of motor proficiency is required to sample information within a workspace – if an individual is unable to move their arm through the space then they will be unable to even begin the learning process. This may explain why assistive forces have been found to help individuals with severe neurological impairment (Lum et al., 2002; Cesqui et al., 2008; Snapp-Childs et al., 2013) or lesser skilled individuals (Sigrist et al., 2013;

Bouchard et al., 2015) – as these systems allow the individual to sample the requisite information and thereby start the learning process.

Our approach is based on the idea that skilful arm movements require accurate predictions about the forces acting on the arm as it moves around the workspace. If these predictions are inaccurate then the system must contend with unexpected perturbations that will force the arm away from its desired trajectory. It has been shown that participants can learn to attenuate the impact of an unexpected perturbation in the short term by developing a ‘global impedance’ strategy, where joint stiffness rapidly increases in response to the application of a sudden unexpected force (Burdet et al., 2001; Burdet et al., 2006). The development of a ‘global impedance’ strategy is a useful short term response to environments which contain unpredictable forces. Nevertheless, skilled continuous movements through a workspace require accurate forward models that allow low entropy, suggesting that the system will seek to learn (and thus predict) the underlying force field in which it is operating. On this basis, we predicted that exposure to a complex force field would, over a sufficient period, drive the system to learn how to move skilfully through the workspace (rather than adopting a short-term impedance strategy).

To test these ideas, we created a metric that quantified the information sampled as individuals learned to move their hand around an artificial environment containing a complex force field (“workspace”). The environment was designed to produce sufficient novelty to limit the possibilities of existing forward models being adapted, but was simple enough that the information acquisition of the exploration of this workspace was able to be modelled. These steps allowed us to examine novel motor learning in two previously conducted experiments whilst providing distinct types of assistive and disturbance forces using an admittance-controlled robotic device. In the second experiment, a condition was created that would enhance learning if the proposed model has merit but would not be expected to benefit learning if the system were simply adopting a short-term global impedance strategy to cope with the force field.

In these experiments, participants had to make continual movements through a workspace comprising a completely novel force field. This arrangement meant that participants had to predict the effects of the underlying structure of the force field – the experiments were not about the participants moving normally and then suddenly experiencing a perturbation of an unpredictable nature. Second, these experiments included baseline measurements of how well the participants could move their arms inside this novel force field. These measurements were taken before and after the participants were given the opportunity to learn the task. The baseline measures did not involve the experimental manipulations (where the robot provided assistive or disruptive forces during the learning process). Thus, the baseline measures provided an index of the motor learning that occurred throughout the experimental sessions. These measures provided the data needed to test the predictions of our new model.

2.3. Procedure

Participants stood in front of a HapticMASTER robotic system (Linde et al., 2002) with a monitor positioned 1.5m away at eye level (Figure 2.1a). The position of the end-effector of the robot was directly mapped (2D only – axes Y and Z) to an on-screen cursor, which updated at 60Hz. A target moved around the screen along a pentagram-shaped trajectory; the participant was told to keep their cursor as close to the target as possible at all times (Figure 2.1d). A single trial consisted of a complete traversal of the pentagram trajectory, split into 5 sub-components (straight lines). The target waited at the end of each sub-component until the participant moved close to the target, then it began moving again. The manipulation of the device of the was made more difficult by a superimposed ‘force field’ workspace, which exerts a force vector on the user’s hand as they move around the workspace based solely on workspace position (Figure 2.1c). Participants attended five 15-minute sessions over a week (one per weekday). Sessions 1 and 5 were pre- and post- tests respectively, and sessions 2-4 were training sessions. The training sessions differed from the pre- and post- sessions in that the target traversed along a vertically flipped pentagram trajectory (Figure 2.1b), and an additional force intervention was applied to the participant’s hand based on allocated group (see section: Groups).

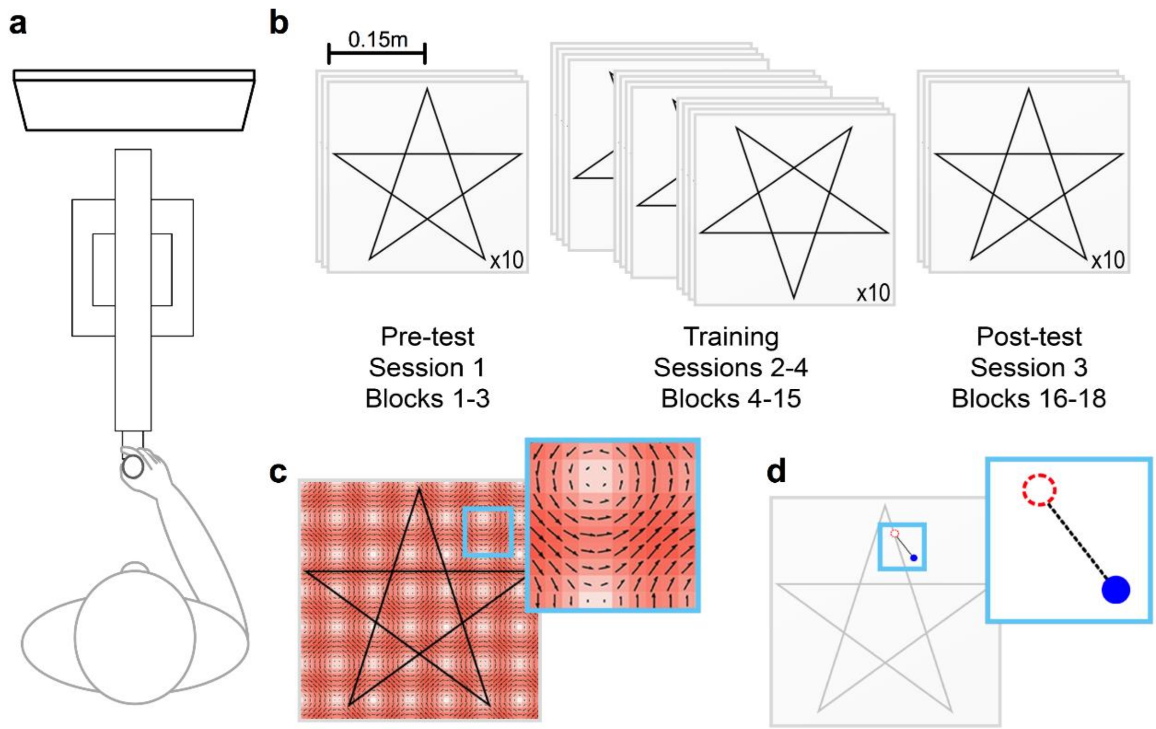


Figure 2.1 – Experiment Design (a) Plan view of the experimental setup showing the relative positions of the participant (bottom), haptic robot arm (middle) and monitor (top); (b) The target trajectories across sessions. The pre- and post-training sessions comprised 3 blocks of 10 trials following a pentagram trajectory (with no error manipulation forces). Training (across three sessions with 4 blocks of 10 trials) included error manipulation forces whilst participants navigated across a vertically rotated pentagram trajectory. (c) Quiver plot of the constant novel workspace force field used across every trial in every condition. Inset shows magnified section (approximate size 5cm x 5cm). Arrows indicate the direction and proportional magnitude of the force vector at discrete locations within the workspace. Relative magnitude is shown from white (no force) through to red (high force). (d) Blue cursor indicates the cursor (hand) position during a trial, the red circle indicates the target, the dotted black line shows the participant's current positional error. Trajectory path and workspace force field remained invisible to participants throughout the experiment.

2.4. Groups

The nature of the force intervention used in the training sessions was varied between groups. This was done by modifying the parameters of the spring in a virtual mass-spring-damper system which was simulated in the HapticMASTER's dedicated haptic rendering computer, which resulted in a force vector being applied to the hand (in addition to the underlying force field). A positive or negative value of k (stiffness) would produce a force towards or away from the target (proportional to distance) respectively. A positive stiffness has the effect of constraining errors, making it easier to stay close to the target, whereas a negative stiffness amplifies errors by pushing the hand away from the target.

2.4.1. Experiment 1

48 right-handed participants (26 male, 22 female) (mean = 29.4 years old, SD = 9.34 years, range 20–59 years). These were randomly allocated in to one of three groups, which used a constant value of k for all training sessions:

- **Assistance** (n = 15): $k = 100\ N/m$, creating a force that pulls the cursor towards the target location.
- **Active-Control** (n = 16): $k = 0\ N/m$, no stiffness intervention.
- **Disruption** (n = 17): $k = -100\ N/m$, creating a force that pushes the cursor away from the target location.

2.4.2. Experiment 2

46 right-handed participants (25 male, 21 female) (mean = 24.93 years old, SD = 6.36 years, range 19–56 years) took part in this experiment. The participants were randomly allocated in to one of three groups, which used different algorithms to select a value of k at the start of each trial in all training sessions:

- **Adaptive Algorithm (AA)** (n = 13): k adjusted each trial based on performance.

- **Adaptive-Disruptive Algorithm (ADA)** (n = 17): k adjusted each trial based on performance, but only decreases (reducing assistance / increasing disruption).
- **Random Algorithm (RAN)** (n = 16): k selected from a uniform distribution $\mathcal{U}(-100, 100)$ at the start of each trial.

Adaptive algorithm stiffness adjustment

In the AA and ADA conditions, the stiffness k was adjusted based on performance. Details of this algorithm are outlined in Jamieson, (2015). Specifically, the authors of this algorithm write:

$$k_{i+1} = f . k_i - g(x_i - x_d) \qquad (1)$$

> *The stiffness, k, of the force field for the next trial is a function of the stiffness in the current trial, i, multiplied by a 'forgetting factor', f, and the difference between the demand error and actual error (x_d and x_n, respectively), multiplied by a gain value, g. The values of f and g dictate the relative sensitivity of the algorithm to previous performance (captured by k_i) and error. The sensitivity of the controller to performances obtained in previous trials is controlled by adjusting f: a larger forgetting factor will weight previous trials more heavily, whereas a smaller forgetting factor will result in more influence by the current trial's force field magnitude. A value of 0.5 was used for both f and g, meaning that half of the weight was made of previous performance and*

> *the other half was made up of the current stiffness setting. This acted to give an equal balance between performance in previous trials, and that in the current trial (Jamieson, 2015, p.132).*

In the ADA algorithm, the force change between trials was lower clamped at 0, meaning only increases in stiffness were allowed.

2.5. Assessment

For the purposes of this experiment, learning was quantified as the decrease in mean path error (absolute distance from cursor to the trajectory) between the pre- and post- test sessions. As the training sessions used a vertically flipped trajectory, this measure of learning is related to the participants' ability to transfer the learning of one trajectory to another. Crucially, the superimposed force field (Figure 2.1c) remains constant throughout the experiment, and so to minimise error in the task the participant presumably must be able to predict and counteract the force field. The experiments can therefore be thought of as a test of how haptic assistance/disruption impacts the learning to track a target under a novel force field.

One-way between subject ANOVAs were performed to examine differences between the groups for the learning measure, and Tukey's post-hoc comparison corrected p values are reported where relevant. Partial eta squared ($\eta^2{}_p$) values are reported for effect size. All data met assumptions of normality through assessment by histogram, Q-Q plots, and Shapiro-Wilk tests.

2.6. Quantifying Information

The underlying force field workspace is the external system the participant must learn in order to perform well in this task. Here, a model is built which quantifies to what extent the participant is exposed to this workspace. The workspace is assumed to be made up of discrete, independent voxels of 1 cm x 1 cm (see Figure 2.2a; total size 40 cm x 40 cm). For the purposes of analysis, this model that assumed participants acquire information about voxels discretely, and any information acquired when the cursor was located inside a particular voxel was 'assigned' to that voxel. The size of 1cm x 1cm voxels was selected as it struck a balance between being too fine grained and too coarse. To ensure that these results were not influenced by this decision, multiple values in orders of magnitudes above and below this value were tested and it was confirmed that they showed the same qualitative pattern of results.

Participants were not explicitly informed about the underlying workspace force field and it remained invisible throughout the experiment. Thus, without the presence of visual information, we assumed that the sensorimotor system would have no reason to predict a change in force as a function of cursor position (at least at the outset of training). This assumption leads to a context where the magnitude of the change in force due to the workspace force field at that point in time corresponds to the force prediction error (i.e. the difference between the experienced and predicted force). As such, new information presented about an individual voxel was equivalent to the change in force at a point in time for the voxel at the cursor position (Figure 2.2b). The force from the workspace force field was a function of position only (a Butterworth filter [cut-off 250Hz] was applied to remove noise).

Information was assumed to be continually acquired at a fixed rate (here, 1000 Hz), and that new information becomes less valuable as a function of the amount of information already acquired about an individual voxel as learning occurs. A parsimonious method by approximating the value of new information with a weighting function was used here- scaling the amount of information presented to an associated information ‘value’.

This function has the desired effect for scaling information – the gradient of the weighting function $= 1$ when information $= 0$ and gradually decreases as new information becomes less valuable. Weighting the information in this way ensures that initial inaccurate estimates about the expected change in force results in high amounts of new information and, as more information is acquired, the value of the new information is lower. The weighting formula, as a function of information presented, was:

$$w(I) = \frac{1}{\lambda} \cdot \log(\lambda I + 1) \tag{2}$$

where log is the natural logarithm and λ corresponds to the weighting. Higher values of λ lead to lower values of information relative to the amount of cumulative information presented, and thus faster learning about a voxel. The reported results have the value of $\lambda = 0.05$, but we tested the model under different assumptions of λ (through values ranging from 0.01 to 1.00) and the pattern remained consistent.

The cumulative (value weighted) information (I_v) related to a particular voxel (i,j) acquired throughout training up to time T, was:

$$I_v(t = T, i, j) = w\left(\int_0^T \Delta f(t, i, j)dt\right) \tag{3}$$

under the assumption that information presented for a particular voxel is the change in force numerically integrated over time for all points in time where the cursor position was inside that voxel (Figure 2.2b).

We also assumed that the total value weighted information acquired was equal to the sum of the value weighted information received from each voxel of the workspace. If the workspace consists of N_x cells horizontally, and N_y cells vertically, the information value for the whole workspace at time T can be calculated as:

$$I_{vT}(t = T) = \sum_{i=0}^{N_x}\sum_{j=0}^{N_y} I_v(t = T, i, j) \tag{4}$$

The total value weighted information assumed that information sampling starts ($t = 0$) at the beginning of the first training session (Session = 2) and completes ($t = T$) at the end of the last training session (Session = 4).

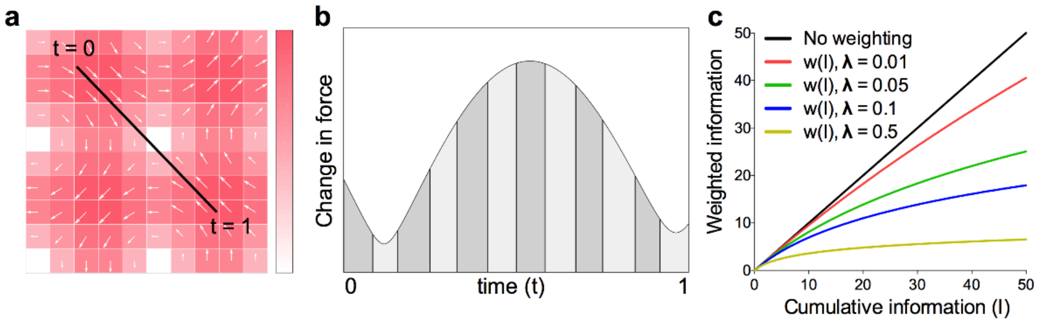


Figure 2.2 – Information quantification (a) Example simulated cursor movement across a sub-section of the workspace (10cm x 10cm). Workspace force field shown as a quiver plot, where higher force magnitude is represented by darker red shading and arrow size, and force direction indicated by arrow orientation. Workspace separated into 1cm x 1cm voxels. (b) Magnitude of change in force measured when moving along the path shown in (*a*) at a constant velocity over 1 second. Vertical black lines indicate the voxel boundary. Shaded regions under the curve separated by the vertical lines represent the information presented which is attributed to the current voxel. (c) Graphical representation of the weighting function for different values of lambda. Note that at higher values of information, the weighted information becomes relatively lower. Note: data on panel (c) is representative of a single voxel and is not related to (a) or (b).

2.7. Results

The amount of motor learning was quantified as the error improvement between the mean pre- and post- path error score (both of which were performed without any stiffness intervention [k = 0] and with the upright pentagram shape). We found significant differences in the amount of motor learning between groups (F (2, 44) = 5.655, p = .0065, η^2_p = .204). Specifically, the group exposed to Disturbance forces during training on the inverted pentagram trajectory had improved significantly more than the Assistance (p = .0136) and the Active-Control (p = .0202) groups (Figure 2.3a). These results are consistent with existing literature in that disruptive forces facilitate greater learning outcomes (Heuer and Lüttgen, 2015).

Group differences in the amount of motor learning from pre- to post-training with no stiffness intervention across the three conditions in Experiment 2 were found

(F (2, 42) = 4.541, p = .0164, η^2_p = .178). There was no statistically reliable difference in learning between the Adaptive Algorithm and Adaptive-Disturbance Algorithm (p = .914). Instead, this effect was driven by improvements following exposure to Random levels of assistance/disruption relative to the Adaptive (p = .018) and Adaptive- Disturbance (p = .009) algorithms (Figure 2.3b).

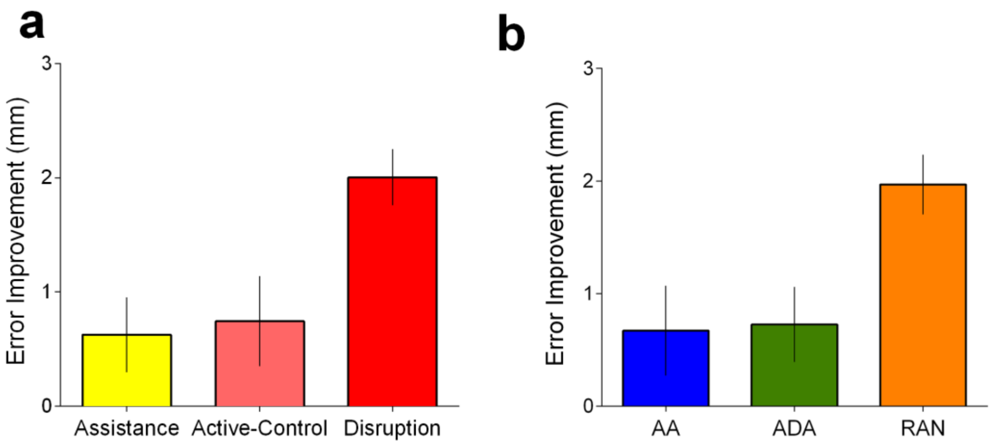


Figure 2.3 – The effects of force interventions on learning. Y-axis shows path error decrease after the training under the force intervention, relative to baseline scores assessed without the force intervention. (a) The disruptive force condition showed enhanced learning relative to the assistive and active-control conditions. (b) Applying a randomly selected level of assistance/disruption (RAN) facilitated more learning compared to algorithms that adapt the assistance/disruption based on performance (AA & ADA).

Data were pooled across both experiments (n = 86) and performed a simple linear regression to predict learning based on cumulative information exposure during training. Consistent with the hypothesis that information exposure predicts performance improvement, a statistically significant relationship was found (F (1, 82) = 10.45, p = .0011), with the information metric explaining 11.2% in variation in learning across all conditions (R^2 = 0.112; Figure 2.4).

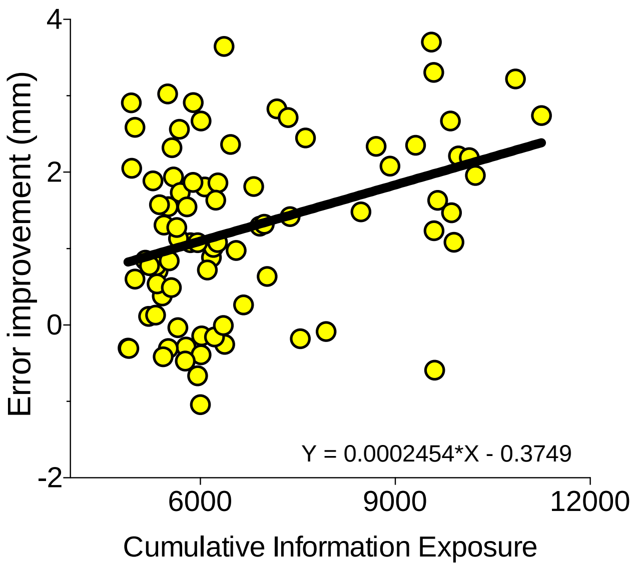


Figure 2.4 – Information exposure predicts learning. Learning (mean path error reduction between pre- and post- training) as a function of cumulative information (arbitrary units) exposure (acquired during training), for all participants in both experiments ($R^2 = 0.122$).

2.8. Discussion

To date, there have been no principled explanations as to why motor learning can be impaired by haptic assistance and facilitated by haptic disruption (Heuer and Lüttgen, 2015). The analysis here uses secondary data to investigate the hypothesis that states that assistive and disruptive forces hinder or facilitate (respectively) the exploration of the dynamics of the task at hand.

To provide a principled account of this explanation, we created a model that quantified the amount of information available to learners during a task. Experiment 1 showed that disturbance forces led to the accumulation of significantly more information across the training period. These results aligned with our analysis of the amount of motor learning following training, whereby the group that sampled more information showed superior performance relative to a group provided with assistance and to an active-control group. In Experiment

2, we demonstrated that the manipulation of information (created by training individuals on a series of random assistive and disturbance forces) yielded better learning compared to providing predictable levels of assistance/ disturbance tuned to individual performance.

Our findings are consistent with previous results suggesting that disturbance forces might be beneficial for motor learning (Emken and Reinkensmeyer, 2005; Cesqui et al., 2008; Lee and Choi, 2010). Importantly, the current work advances these reports by providing, and testing, a theoretical account of why disruptive forces might accelerate learning. Specifically, we show that these results can be predicted by an information theory-based account of parameter exploration in motor learning. Here, motor learning is seen as a process of uncertainty reduction through development of forward models that make better predictions (Wolpert and Miall, 1996; Kawato and Wolpert, 2007). The decrease in uncertainty relates to improved inferences created by the system through exposure to information that relates perceptual output to motor signal input.

In line with this explanation, through pooling the data across both experiments, we found that the amount of workspace information that participants were exposed to during training could predict a statistically significant amount of variance in learning. Whilst this is a relatively small effect, given the plethora of variables that could also have influenced learning across these different manipulations (six experimental conditions in two experiments), it is notable that this relationship between information and learning could be detected.

Our results build on previous work showing a relationship between variability and motor learning. For example, van Beers (2009) showed that the random effects of planning noise accumulate, in contrast to task-relevant errors which show close to zero accumulation (explained by effective trial-by-trial corrections), whilst Wu et al's experiments (2014) (results described earlier), have shown that task-relevant motor variability facilitates faster learning rates. On these grounds, it has been argued that intrinsic movement variability leads to motor exploration, which sub-serves motor learning and performance optimization. Indeed, the idea that action exploration can drive learning has long been mooted in theories of operant behaviour (Barto et al., 1998) and human development (Bruner, 1973; Gibson, 1988; Thelen, 1989). Recent experiments have shown that (a) artificially manipulating the relationship between movements and visuomotor noise can be used to teach people specific control policies (Thorp et al., 2017) and (b) the variability in task-redundant parameters can predict motor adaptation rates (Singh et al., 2016). The current findings demonstrate that extrinsic variability delivered through haptic disturbance can, in the same vein, augment learning by increasing the amount of information sampled by the learner.

The general notion that increased exposure to information can lead to faster learning is well explained by theories of structural learning and has good support from a range of empirical studies (Braun et al., 2009; Johnson et al., 2010; Braun, Mehring, et al., 2010; Braun, Waldert, et al., 2010; Turnham et al., 2011; Yousif and Diedrichsen, 2012) including investigations of laparoscopic surgical training (White et al., 2013). Our extension to these ideas is that learning of the structure can be directly related to the amount of information available to the learner. Indeed, regression analyses for our data shows that the amount of information accumulated over training (as indexed by our model) provided greater explanatory power compared to a measure of motor variability alone in this task.

These findings raise the issue of which neural substrates underpin these learning processes. The neural processes that implement the computational algorithms exploited by the human nervous system remain to be discovered (Wolpert, 1997; Wolpert et al., 2001). Likewise, the underlying control mechanisms supporting skilled arm movements are poorly understood and, as such, it is difficult to speculate on how the individuals learned to compensate for the complex force field, but we suggest that the learning was likely to involve processes related to optimal feedback control as well as predictive mechanisms (Todorov and Jordan, 2002; Todorov, 2004; Franklin and Wolpert, 2011).

Our findings suggest that the participants developed forward or inverse models that allowed them to predict (and thus compensate for) the novel force field through which they needed to move. It has been shown previously that participants can learn a short term strategy of stiffening their arm to resist the effects of sudden unexpected force perturbations (Burdet et al., 2001; Burdet et al., 2006). This work has demonstrated that humans learn to use selective control of impedance geometry in order to stabilise unstable dynamics in a skilful and energy efficient manner. It is probable that participants in the current experiments adopted such a strategy when they were first exposed to the novel workspace (as they were unable to predict the forces that were applied as they moved through the space). Importantly, there was a regular (lawful) structure to the novel workspace, in the same way that the world provides a lawful force field through which the neonate must learn to move their arm. We hypothesised that the system would learn the underlying force field so that the arm could move skilfully through the workspace rather than repeatedly contend with unexpected displacement. Experiment 2 allowed us to test whether participants were learning the force field or adopting a global impedance strategy, by which the arm is stiffened in all directions to counteract external force interventions. As outlined above and demonstrated in previous research, participants are likely to adopt a global impedance strategy when the force intervention is largely disruptive and increases error ($k < 0$). However, in Experiment 2, the random condition consisted of (on average) 50% assistive trials, whereby the force intervention *assisted* movement, thus rendering such a strategy sub-optimal. We reasoned that, in contrast to the random forces, the adaptive disruption algorithm, where participants were provided with a more consistent presentation of disturbance forces would be more likely to adopt an impedance control strategy. Given that we observed improved learning in the random condition, impedance control is unlikely to provide a full account of these data. Instead, these results indicate that participants were learning to skilfully counteract the underlying workspace force field and we propose that this learning was promoted, in part, through the increased information acquired during training.

It is important to note that this study used neurologically intact adults as participants and whilst the force field in the two experiments allowed us to examine novel skill learning, the difficulty was tuned to a level such that all participants could complete the task. We speculate that disrupting the training of individuals with neurological deficits (e.g. cerebral palsy) might not be beneficial, and constraining errors in these populations could speed up learning by helping the individuals sample the necessary information (Snapp-Childs et al., 2013). Consistent with this, there is work with stroke survivors that has shown that error amplification is useful in rehabilitation for mild impairment, but error guidance is necessary for patients with more severe damage (Marchal-Crespo and Reinkensmeyer, 2009). Likewise, haptic guidance has been found to be beneficial for people with relatively low skill levels, but error enhancement is better for highly skilled individuals (Milot et al., 2010; Sigrist et al., 2013; Bouchard et al., 2015). The current work builds on these observations and provides a theoretical framework for the development of optimized robotic training devices in skill training and rehabilitation.

Finally, we note that these finding do not imply a direct causal relationship between exploration of task dynamics and motor learning. Instead, manipulation of the task dynamics through means other than through a secondary force intervention might provide supporting evidence for such a relationship and we explore this topic further in the subsequent chapter.

2.9. Acknowledgements

The experimental task was originally presented in Jamieson (2015). Since this report, additional data were collected (Experiment 1, N = 11; Experiment 2, N =

8) to increase statistical power. All subsequent analyses and model developments were undertaken by the present author.

3. Direct Manipulation of Information Acquisition in Motor Learning

3.1. Overview

The modelling analysis of secondary data presented in the previous chapter indicates a relationship between information acquisition and motor learning. Specifically, we proposed that an information-theoretic account could reconcile the seemingly paradoxical findings that increasing disturbances in force field learning tasks can accelerate learning and we tested these ideas with data from two experiments. The implications of this idea are that disruptive forces per se may not be the driving force behind learning, but that learning arises from a by-product of these conditions- allowing individuals to sample more of the task space, thus acquiring information that can subsequently be utilised to improve performance. Should these ideas hold, we expect a motor learning task that increases information sampling to lead to superior learning than one that constrains information in the absence of any perturbations. We test this idea in the present chapter.

3.2. Introduction

Th ability for humans to adapt to the dynamics of a task is a fundamental aspect of learning. As we move, forces are generated by the tools and objects we interact with, as well as the weight of our own bodies. These forces must be compensated for by equal and opposite forces to stabilise motion. It is thought that these forces are mitigated by learning how to predict them (Emken and

Reinkensmeyer, 2005), as internal forward models of the task dynamics are formed (Wolpert et al., 2011). In this way, it has been reasoned that exposing participants to more of the properties of the task could accelerate motor learning (Emken and Reinkensmeyer, 2005; Braun et al., 2009). This possibility has some credibility following evidence suggesting disruptive forces accelerate learning (Sigrist et al., 2013; L. Marchal-Crespo et al., 2014; Heuer and Lüttgen, 2015), and the previously unexplored mechanistic explanation that implies accelerated learning is due to greater exploration and exposure to the task dynamics- a common side effect of forces that push users away from a target.

In Chapter 2, we showed a relationship between workspace information acquisition and motor learning in a target-tracking task completed under a force field. The task was designed to require development of forward models that can predict (and therefore mitigate) the effects of the workspace force-field. Training was performed with error-augmenting forces, and forces that increase error (disruptive) were found to facilitate learning. In a second experiment, the level of assistance/disruption was tuned to performance, but counterintuitively, a condition that provided random levels of assistance/disruption on a trial-by-trial basis facilitated enhanced learning.

To explain these results, we proposed that the amount of workspace information that the participant acquired during training could account for the exhibited improvements in learning. To provide a more robust test of this putative relationship between information and motor learning, here we introduce a new experiment designed to facilitate the acquisition of information without the use of disruptive forces. Specifically, we directly manipulated the target trajectory to modify the amount of information participant were exposed to during training. Participants were randomly allocated to one of two conditions- a ‘High Variability’ group, where position around the workspace was varied to a large degree (and thus allowed participants to sample more of the workspace) and a ‘Low Variability’ group, where the position was relatively stable and exposure to the total workspace was limited. If the predictions formed in the previous chapter hold, we should expect to see participants exposed to more information during training (the High Variability group) to have the larger improvements from baseline to post-training compared to the Low Variability group.

3.3. Procedure

The experimental procedure was identical to Experiment 1 & 2 in Chapter 2. However, the stiffness of the spring in the virtual mass-spring-damper system was set to $k = 0$ to remove any haptic guidance or disruption. Instead, only the target trajectory was modified between groups to manipulate information exposure. The same background force field was used for all trials across the experiment. Pre- and post- tests were 30 trials with A training session was comprised of six trials with two 30 second rests.

3.4. Groups

In this experiment, there were two different training trajectories, designed to manipulate the amount of information about the force field the participant is exposed to while controlling for other variables. This was done with a factor trajectory variability (TV) with two factors ‘High Variability’ (HV) and ‘Low Variability’ (LV), which were intended to vary the participants’ position around the workspace a high and low amount, respectively. The trajectories were based on the inverted pentagram used previously but were effectively shifted around the workspace in five possible positions. This was done so that the general type of motor task is the same as in previous experiments (i.e., movements of approximately 30cm in various directions), and that the total path length (and therefore, time per trial) was virtually fixed between groups (this is verified in Results, Figure 3.2b). The points on the trajectories were selected to utilise the largest amount of the usable workspace of the device.

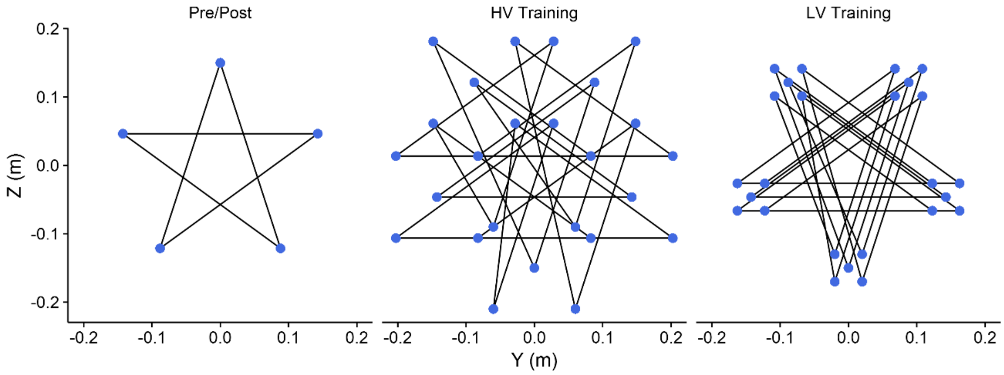


Figure 3.1 – Trajectories used in Experiment 2. Total path lengths: Pre/post = 1.43m (x5 = 7.13m), HV = 7.18m, LV = 7.14m. The high variability training section therefore required only around 4cm (0.5%) extra movement per trial.

A second factor in the experiment was the hand used (HU), i.e. the use of preferred on non-preferred hand. The hand manipulation was used to test the hypothesis that performance increases would be greater in participants using non-preferred hand. The 2x2 design of the experiment is shown in Table 1.

Table 1 – 2x2 between-subjects design for Experiment 2

		Trajectory variability (TV)	
		High (HV)	Low (LV)
Hand used (HU)	Preferred	HV + P $n = 10$	LV + P $n = 10$
	Non-preferred	HV + NP $n = 8$	LV + NP $n = 8$

3.5. Results

Four participants' data were excluded from analyses due to high variances in error between sessions. A two-way ANOVA was conducted to compare the effects of TV and HU, and the interaction between TV and HU on the amount of information acquisition during training. Effect size (generalised eta squared; η^2_p) is reported. All data met assumptions of normality through assessment by histogram, Q-Q plots, and Shapiro-Wilk tests. There was a statistically significant main effect of TV (F(1, 32) = 8.540, p = .006; η^2_p = .211; Figure 3.2a). The main effect of HU yielded a non-significant result (F(1, 32) = 1.479, p = .232, η^2_p = .044). The interaction between TV and HU was also non-significant (F(1, 32) = 0.358, p = .554, η^2_p = .011).

A t-test was performed comparing TV group as a predictor of total training time, as a test of whether the different trajectories significantly affected the time spent training. *Cohen's d* effect size is reported. No significant difference in training time was found (t(34) = -0.904, p = .372, *d* = 0.976; Figure 3.2b).

To test if the use of non-preferred hand affected performance improvement, a t-test comparing NP vs P participants in terms of performance improvement (reduction in average path error pre- to post-test). The hand used significantly affected performance improvement (t(34), 3.1713, p = .003, *d* = 1.064), with NP and P having mean error reduction scores of 2.806mm and 0.405mm respectively.

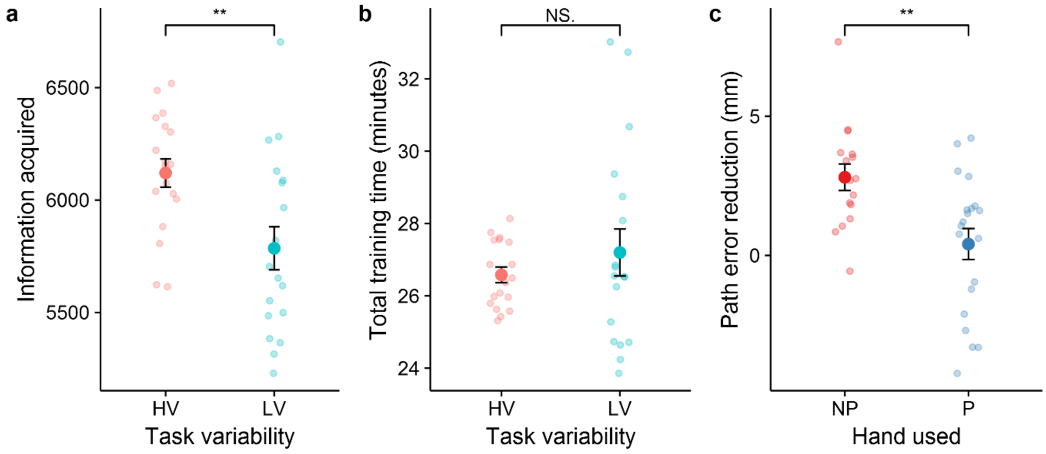


Figure 3.2 – Verification of experimental design. (a) The high variability group were exposed to significantly more information than the low variability group, after analysing the training data with the information model. (b) There was no significant difference in training time (cumulative movement time for all training trials) between TV. (c) The use of preferred or non-preferred hand significantly affected performance improvement. Error bars represent +/-1 standard error of the mean.

To test the hypothesis that increased information exposure results in a greater increase in performance, a two-way ANOVA was conducted to examine the effects of trajectory variability (TV) and hand used (HU), and the interaction between TV and HU, on the increase in performance between pre- and post-tests. There was no significant main effect of TV on performance (F(1, 32) = 0.007, p = .932, η^2_p < .001). Hand used significantly affected performance improvement (F(1, 32) = 9.975, p = .003, η^2_p = .238), but there was no significant interaction effect between TV and HU on performance improvement (F(1, 32) = 1.717, p = .199, η^2_p = .051; Figure 3.3b).

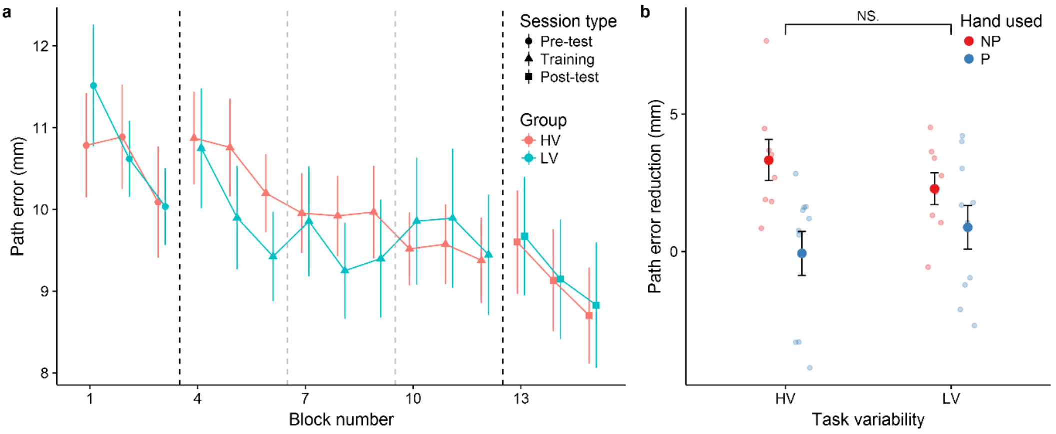


Figure 3.3 – Performance improvement. (a) Participants reduced their error over time. Dashed vertical lines separate sessions (days). (b) Path error improvement was not significantly different between TV groups. Error bars represent +/-1 standard error of the mean.

A linear regression was also performed to predict performance improvement based on information acquisition, irrespective of group. The relationship between path error improvement and information was not statistically significant (F(1, 34) < 0.001, p = .987, R^2 = -0.029; Figure 3.4).

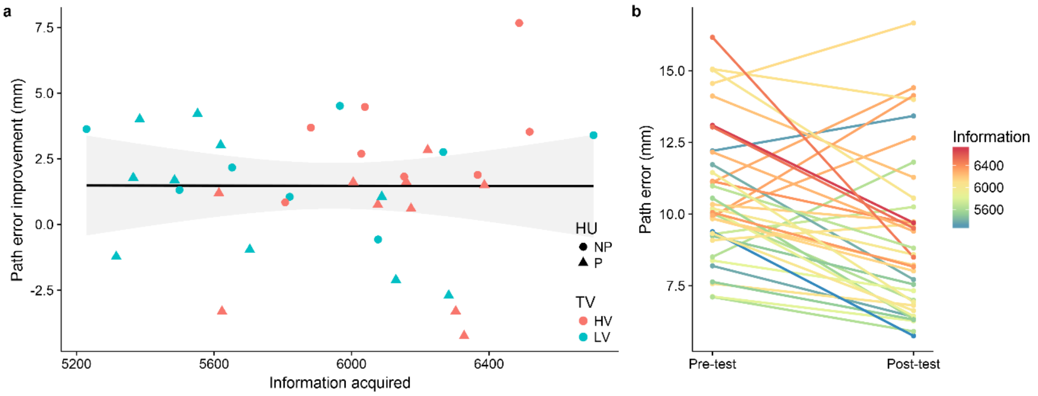


Figure 3.4 – Information acquisition does not predict performance improvement. (a) There is no correlation between information acquired and path error improvement. (b) Visualisation of performance improvement for individual participants between pre- and post-tests for differing levels of information acquisition.

3.6. Discussion

This experiment aimed to investigate the hypothesis that information acquisition through means of workspace exploration could accelerate the learning of a novel force-field. The target trajectory of a moving target was modified between groups to manipulate the amount of information participants were exposed to. A mathematical model of information acquisition developed in Chapter 2 was applied to the movement data collected from this new experiment.

The trajectory variability (TV) was found to have a significant effect on the amount of information acquired as calculated by the model. Furthermore, there were no significant differences in training time, indicating that the trajectory variability manipulated information without changing the total training time. This indicates that the experimental manipulation worked as intended. However, the results did not match those predicted by the hypothesis. The amount of information exposure was hypothesised to correlate with performance improvement, but we found no relationship here. Instead, the only differences in performance improvement were observed in the hand used factor, where improvement was significantly higher in those who used their non-preferred hand vs those using preferred hand.

There are several ways of interpreting the results. First, it may be the case that the information model is an inaccurate method of quantifying information though means of workspace exploration. The model is built on several assumptions: One, that information acquisition is proportional to a change in force. This would mean that areas of the workspace which have a greater force-derivative would produce expose more information to the participant (see Figure 2.2b). This seems intuitive, as from this emerges the property that a more complex (high entropy) force field with large changes in force would hold more information which can be subsequently acquired by the user and aligns with the Shannon view of information (Shannon, 1948). However, since a change in force only occurs when moving, this assumption would mean information under this model can be acquired just by increasing average speed (or movement path length). Informal parameter exploration reveals however that removing this aspect of the model favouring a constant information acquisition rate did not alter the pattern of results (data not shown). A second assumption for the information model is that learning about the workspace is done in a spatially-discrete manner. For example, this would mean that acquiring information about a sub-section of the workspace would not give participants any information about other areas of the workspace, i.e. there is no generalisation.

This artefact could be responsible for the large difference in information between TV groups in this experiment, since the HV group were exposed to a larger number of these voxels (sub-sections) for any given trial. The weighting function punishes repeated acquisition of information from a small number of voxels and rewards exploration. However, here we are assuming information acquisition happens per voxel of the workspace and the only way to refine the model is to physically re-enter that area of the workspace. This is particularly problematic since the nature of the workspace is that it has some structure and repetition (Figure 2.1C); the learning of the structural parameters of the workspace was not modelled here, and assumes more of a model-free mechanism of learning. A non-repeating pattern with more unpredictable forces would mitigate some of the problems that arise from this assumption, or a better model formulation that estimates the rate of model-based learning of the structure of the workspace forces.

A second way of interpreting the results shown, is that this model works as intended (captures information acquired through workspace exploration), but that the assumption that learning arises from workspace exploration is incorrect. One aspect that the assumption of workspace exploration facilitating information acquisition lacks is that learning is a passive process of observation. Workspace exploration from a purely observational point of view does not consider participant’s actions on the device in order to mitigate the effects of the force field. There are also other parameters which would affect the building of a model that can counteract the workspace, for example learning to mitigate the force field at a range of velocities, from different directions, and with a different target position (the task is to stay close to a target position). While this result shows manipulation of exploration of the physical workspace, exploration of these other parameters was not manipulated. For example, we did not manipulate error rates in this experiment and it is clear that they are key part of the learning process (Laura Marchal-Crespo et al., 2014).

Because previous results have shown consistently that haptic disruption enhances motor learning in skilled subjects, there are several other models which may be able to explain learning advantages for haptic disruption, while accounting for the apparent lack of advantage given by purely increased workspace exploration in this task. It has been previously observed that the use of repelling force fields can increase limb stiffness in participants by co-contracting muscles in the limb (Osu et al., 2002; Franklin et al., 2003; Heuer and Lüttgen, 2015), which in-turn reduces error. However, this hypothesis can exist simultaneously with one based on information: If co-contracting of muscles is a learned behaviour which is facilitated more by error amplifying forces (compared with error reducing or no guidance), the feedback received throughout training (visual, kinaesthetic, proprioceptive) must have somehow led the motor system to construct the model which outputs a co-contraction strategy (even when no error amplifying force is present).

Another hypothesis is that disruptive forces facilitate ‘error-based learning’ though constant error amplification and requirement for the participant to refine their on-line control to continually correct for these errors (Milot et al., 2010; Wolpert et al., 2011; Laura Marchal-Crespo et al., 2014). Similarly, under the error-based learning hypothesis the informational content of the feedback interpreted by the motor system has been augmented (due to the haptic error amplification) such that it speeds up learning and can perform well even when the error amplifying force is removed. Or perhaps under disruptive forces, error-correcting actions are required to be either more frequent, speeding up the formation of internal models; or more intense, perhaps meaning the feedback has a greater signal-to-noise ratio.

3.7. Conclusions

The claim that information acquisition through workspace exploration drives learning in this task of tracking a moving target in a novel force field is not supported by these data. Rather than using forces to facilitate exploration of the workspace, the target trajectory was modified such that exploration became an explicit part of training. Under the information model presented in Chapter 2, information acquisition was found to be greater for the HV group compared to the LV group. This was the case, but there were no significant differences in performance improvement found, implying information acquisition (under this model) has no causal relationship with learning. However, there are other features of the task which provide the participant with information which were not included in the model, most notably, the positional error. This experiment purposely did not attempt to modify the informational content of the positional error feedback, in order to solely investigate the workspace exploration aspect of the task. Subsequent studies should investigate the role that the positional error feedback plays in refining our ability to perform accurate on-line corrections, and if it is possible to exploit these mechanisms to improve performance in real world tasks.

4. Using Virtual Reality to Study Human Behaviour

4.1. Overview

Virtual Reality (VR) systems offer a powerful tool for human behaviour research. The ability to create three-dimensional scenes and measure responses to the visual stimuli enables the behavioural researcher to test hypotheses in a manner and scale that were previously unfeasible. For example, a researcher wanting to understand interceptive timing behaviour might wish to violate Newtonian mechanics, so objects move in novel 3D trajectories. The same researcher may wish to collect such data with hundreds of participants outside the laboratory, and the use of a VR headset makes this a realistic proposition. The difficulty facing the researcher is that sophisticated 3D graphics engines (e.g. Unity) have been created for game designers rather than behavioural scientists. In order to overcome this barrier, we have created a set of tools and programming syntaxes that allow logical encoding of the common experimental features required by the behavioural scientist. The Unity Experiment Framework (UXF) allows the researcher to readily implement several forms of data collection and provides researchers with the ability to easily modify independent variables. UXF does not offer any stimulus presentation features, so the full power of the Unity game engine can be exploited. We use a case study experiment, measuring postural sway in response to an oscillating virtual room, to show how UXF can replicate and advance upon behavioural research paradigms. We show that UXF can simplify and speed up development of VR experiments created in commercial gaming software and facilitate the efficient acquisition of large quantities of behavioural research data. We use this software to develop the experimental tasks reported in subsequent chapters.

4.2. Introduction

Virtual Reality (VR) systems are opening up new opportunities for behavioural research as they allow visual (and auditory) stimuli to be displayed in 3D computer generated environments that can correspond to the participant's normal external Cartesian space, but which do not need to adhere to the rules of Newtonian mechanics (Wann and Mon-Williams, 1996). Moreover, VR systems support naturalistic interactions with virtual objects and can provide precise measures of the kinematics of the movements made by adults and children in response to displayed visual stimuli. In addition, the relatively low cost and portability of these systems lowers the barriers to performing research in non-laboratory settings.

The potential advantages of VR in behavioural research have been recognised for at least two decades (e.g. Loomis, Blascovich, & Beall, 1999) but recent advantages in technology and availability of hardware and software are making VR a feasible tool for all behavioural researchers (rather than a limited number of specialist VR labs). For example, researchers can now access powerful software engines that allow the creation of rich 3D environments. One such popular software engine is Unity (alternatively called Unity3D; Unity Technologies, 2018). Unity is a widely used 3D game engine for developing video games, animations and other 3D applications and it is growing in its ubiquity. It is increasingly being used in research settings as a powerful way of creating 3D environments for a range of applications (e.g. psychology experiments, surgical simulation, rehabilitation systems). The recent popularity of VR head-mounted displays has meant that Unity has become widely used by games developers for the purpose of crating commercial VR content. Unity has well developed systems in place for rich graphics, realistic physics simulation, particles, animations and more. Nevertheless, it does not contain any features specifically designed for the needs of human behaviour researchers. We set out to produce an open source software resource that would empower researchers to exploit the power of Unity for behavioural studies.

A literature search of human behavioural experiments reveals that experiments are often defined by a common model, one that more easily allows researchers to exercise the scientific method. Experiments are often composed of trials, where trials can be defined as an instance of a scenario. Trials are usually composed of a stimulus and a human response and are a basic unit of behavioural experiments. Trials can be repeated many times for a single participant, increasing the signal-to-noise ratio of measurements, or allowing the study of human behaviour over time (e.g. adaptation and learning). Blocks can be defined as a grouping of trials that share something in common; comparing measures between blocks allows the examination of how substantial changes to the scenario affect the response. A session is a single iteration of the task with a participant. Defining an experiment in such a session-block-trial model (Figure 4.1) allows the definition and communication of an experimental design without ambiguity.

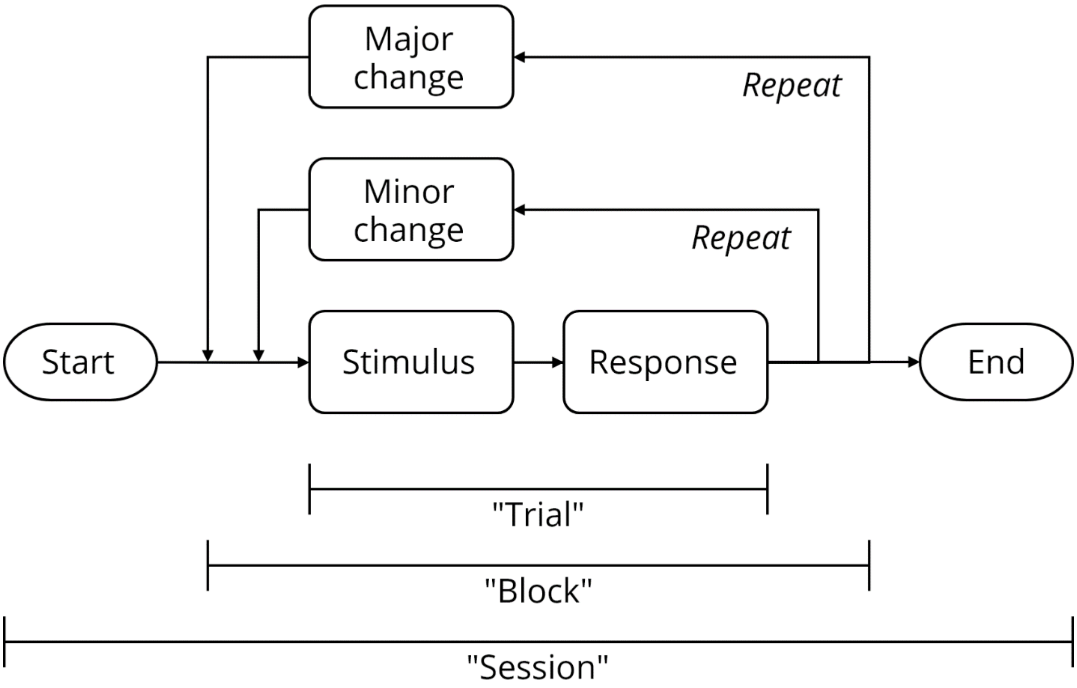


Figure 4.1 – Structure of typical human behaviour experiments, in the session-block-trial model. Many experiments comprise multiple repetitions of trials. Between trials, only minor changes are made. A substantial change of content in the trial is often described as creating a new “block”. A single iteration of a task by a participant is called a session.

The use of this session-block-trial model in computer-based experiments affords a certain type of system design structure that mirrors the model itself. Typically, the code produced for an experimental task consists of a loop, where the process of presenting a stimulus and measuring a response is repeated many times, sometimes changing the parameters between loop iterations. The popularity of this experimental architecture means that researchers have attempted to provide tools that allow the development of tasks without the need to ‘reinvent the wheel’. Relatedly, development of the stimuli for software experiments is often difficult without knowledge of low-level computer processes and hardware. Thus, several software packages have been released which aim to make the stimuli themselves easier to specify in code. There is some crossover between these two types of packages, some focus only on stimuli whilst others also provide high-level ways to define the trials and blocks of the experiment and we briefly consider some of the most commonly used tools next.

PsychToolbox (Brainard, 1997) is a software package for MATLAB that allows researchers to program stimuli for vision experiments, providing the capability to perform low-level graphics operations but retaining the simplicity of the high-level interpreted MATLAB language. PsychoPy (Peirce, 2007) is an experimental control system that provides a means of using the Python programming language to systematically display stimuli to a user with precise timing. It consists of a set of common stimulus types, built-in functions for collection and storage of user responses/behaviour, and means of implementing various experimental design techniques (such as parameter staircases). PsychoPy also attempts to make research accessible for non-programmers with its 'builder', a GUI (graphical user interface) that allows development of experiments with little to no computer programming requirements.

The graphics processes for immersive technologies are significantly more complex than those required for two dimensional displays. In VR, it is difficult to think of stimuli in terms of a series of coloured pixels. The additional complexity includes a need for stimuli to be displayed in apparent 3D to simulate the naturalistic way objects appear to scale, move and warp according to head position. Unity and other game engines have the capacity to implement the complex render pipeline that can accurately display stimuli in a virtual environment; current academic focused visual display projects may not have the resources to keep up with the evolving demands of immersive technology software. Vizard (WorldViz, 2018), Unreal Engine (Epic Games, 2018), and open-source 3D, game engines such as Godot (Godot, 2018) and Xenko (Xenko, 2018) are also feasible alternatives to Unity, but Unity may still be a primary choice for researchers because of its ease of use, maturity, and widespread popularity.

4.3. The Unity Experiment Framework (UXF)

To provide behavioural researchers with the power of Unity and the convenience of programs such as PsychoPy, we created the Unity Experiment Framework (UXF). UXF is a software framework for the development of human behaviour experiments with Unity and the main programming language it uses, C#. UXF takes common programming concepts and features that are widely used, and often re-implemented for each experiment, and implements them in a generic fashion (Table 2). This gives researchers the tools to create their experimental software without the need to re-develop this common set of features. UXF aims to specifically solve this problem, and overtly excludes any kind of stimulus presentation system, with the view that Unity (and its large asset developing community) provides all the necessary means to implement any kind of stimulus or interaction system for an experiment. In summary, UXF provides the ‘nuts and bolts’ that work behind the scenes of an experiment developed within Unity.

Table 2 – Common experiment concepts and features which are represented in UXF

Concept	Description
Trial	The base unit of experiments. A trial is usually a singular attempt at a task by a participant after/during the presentation of a stimulus.
Block	A set of trials – often used to group consecutive trials that share something in common.
Session	A session encapsulates a full "run" of the experiment. Sessions are usually separated by a significant amount of time and could be within subjects (for collection of data from a singular participant over several sessions) and/or between subjects (for collection of data from several participants each carrying out a single session).
Settings	Settings are parameters or variables for an experiment, block, or trial, usually predetermined, that quantitatively define the experiment. Settings are useful for defining the experimental manipulation (i.e. the independent variables).
Behavioural data	We perform an experiment to measure the effect of an independent variable on a dependent variable. Behavioural data collection allows for the collection of measured values of dependent variables on a trial-by-trial basis. For example, we may wish to collect the response to a multiple-choice question, or the distance a user throws a virtual ball.
Continuous data	Within a trial, we may want to measure a value of one or more parameters over time. Most commonly we want to record the position and rotation of an object within each trial. This could be an object that is mapped to a real-world object (e.g. participant head, hands) or a fully virtual object (virtual ball in a throwing experiment). Position and rotation of an object is the main use case but UXF supports measurement of any parameter over time (e.g. pressure applied to a pressure pad).
Participant information	There may be other variables that we cannot control within the software which we may wish to measure to record to examine its relationship to the result. For example, age or gender of the participant.

4.3.1. Experiment structure

UXF provides a set of high-level objects that directly map onto how we describe experiments. The goal is to make the experiment code more readable and avoid the temptation for inelegant if-else statements in the code as the complexity increases. Session, blocks, trials are our ‘objects’ which can be represented within our code. The creation of a session, block or trial automatically generates properties we would expect them to have – for example each block has a block number, each trial has a trial number. These numbers are automatically generated as positive integers based on the order in which they were created. Trials contain functionality such as ‘begin’ and ‘end’ which will perform useful tasks implicitly in the background, such as recording the timestamp when the trial began or ended. Trials and blocks can be created programmatically, meaning UXF can support for any type of experiment structure, including staircase or adaptive procedures.

4.3.2. Measuring dependent variables

While the trial is ongoing, at any point we can add any observations to the results of the trial, which will be added to the behavioural data .CSV output file at the end of the session. Additionally, we can continuously log a variable over time at the same rate as the display refresh frequency (90Hz in most currently-available commercial VR HMDs). The main use case of this is where the position and rotation of any object in Unity can be automatically recorded on a per-trial basis, saving a single .CSV file for each trial of the session. This allows for easy cross-referencing with behavioural data. All data files (behavioural, and continuous) are stored in a directory structure organised by *experiment* > *participant* > *session number*.

4.3.3. Setting independent variables

Settings can be used to attach values of an independent variable to an experiment, session, block, or trial. Settings have a cascading effect, whereby one can apply a setting to the whole session, a block or a single trial. When attempting to access a setting, if it has not been assigned in the trial, it will attempt to access the setting in the block. If it has not been assigned in the block, it will search in the session (Figure 4.2). This allows users to very easily implement features common to experiments, such as “10% of trials contain a different stimulus”. In this case, one could assign a “stimulus” setting for the whole session, but then assign 10% of the trials with a different value for a “stimulus” setting.

Settings are also a useful feature for allowing for changing experimental parameters without modifying the source code. A simple text file (JSON format) can be placed in the experiment directory which will be read upon the start of a session, and its settings applied to that session. This system speeds up the iteration time during the process of designing the experiment; the experimenter can change settings from this file and see their immediate effect without changing any of the code itself. It also allows multiple versions of the same experiment (e.g. different experimental manipulations) to be maintained within a single codebase using multiple settings files. One of these settings profiles can be selected by the experimenter on launching the experiment task.

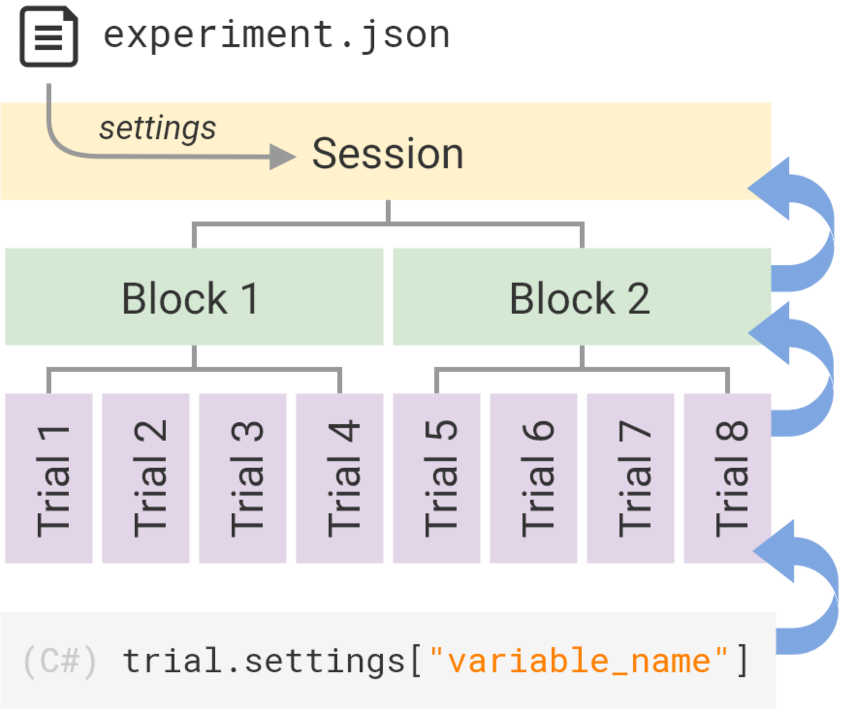
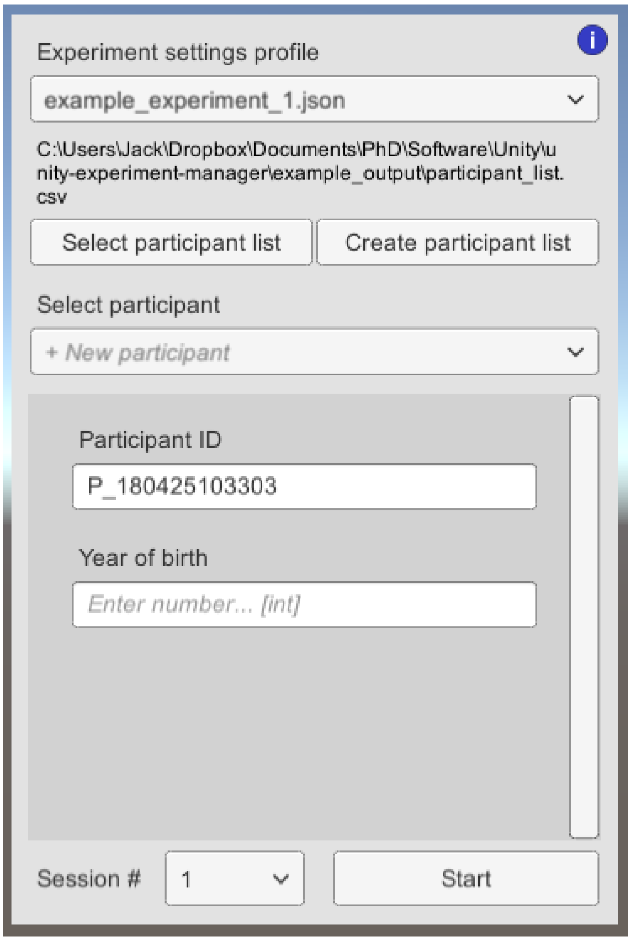


Figure 4.2 – The UXF Settings system. Independent variables that we change in order to iterate a design of an experiment, or to specify the experimental manipulation itself, can be written in a human-readable .json file. Settings can also be programmatically accessed or created at trial, block or session level. Where a setting has not been specified, the request cascades up and searches in the next level above. This allows both “gross” (e.g. to a whole session) or “fine” (e.g. to a single trial) storage of parameters within the same system.

4.3.4. Experimenter User Interface

UXF includes an (optional) experimenter user interface (UI) to allow selection of a settings profile, and inputting additional participant information, such as demographics. Information the experimenter wishes to collect is fully customisable. The UI includes support for a “participant list” system, whereby participant demographic information is stored in its own CSV file. As new participants perform the experiment, their demographic information is stored in the list. This allows participant information to be more easily shared between sessions or even separate experiments – instead of having to input the information each time, the experimenter can select any existing participant found in the participant list via a drop-down menu.



Experiment settings profile

example_experiment_1.json

C:\Users\Jack\Dropbox\Documents\PhD\Software\Unity\unity-experiment-manager\example_output\participant_list.csv

Select participant

+ New participant

Participant ID

P_180425103303

Year of birth

Enter number... [int]

Session #

1

Figure 4.3 – Screenshot of the experimenter user interface.

4.3.5. Example

Below is an example of the C# code used to generate a simple 2 block, 10 trial experiment where the participant is presented with a number x and they must input the doubled value ($2x$).

```
// create variable: block 1, containing 5 trials
```

```
var block1 = session.CreateBlock(5);
// apply a setting ‘manipulation’ as false to the whole block
block1.settings[‘manipulation’] = false;
// loop over the trials and assign the setting ‘x’ to a random value
foreach (var trial in block1.trials)
    trial.settings[‘x’] = Random.Range(1, 10);

// create variable: block 2, containing 5 trials
var block2 = session.CreateBlock(5);
// apply a setting ‘manipulation’ as true for the whole block
block2.settings[‘manipulation’] = true;
// loop over the trials and assign a the setting ‘x’ to a random value
foreach (var trial in block2.trials)
    trial.settings[‘x’] = Random.Range(1, 10);

// apply a setting to only the first trial of block 1
block1.firstTrial.settings[‘show_instructions’] = true;

// begin the first trial...
session.firstTrial.Begin();
```

Elsewhere in our project, we must define what happens when we begin the trial (such as making the value of x appear for the participant), and mechanisms to retrieve the participant’s response for the trial (participant’s calculated value of $2x$). These are to be created with standard Unity features for making objects appear in the scene, collecting user response via keyboard input, etc. The resulting behavioural data .CSV file would be automatically generated and saved (Table 3). A typical structure of a task developed with UXF is shown in Figure 4.4.

Table 3 – Example behavioural data output. Columns not shown include participant ID, session number, and experiment name.

trial_num	block_num	start_time	end_time	manipulation	x	response
1	1	0.000	1.153	FALSE	8	16
2	1	1.153	2.112	FALSE	3	6
3	1	2.112	2.950	FALSE	4	8
4	1	2.950	3.921	FALSE	7	14
5	1	3.921	4.727	FALSE	4	8
6	2	4.727	5.826	TRUE	9	18
7	2	5.826	6.863	TRUE	5	10
8	2	6.863	7.693	TRUE	10	20
9	2	7.693	8.839	TRUE	6	12
10	2	8.839	9.992	TRUE	3	6

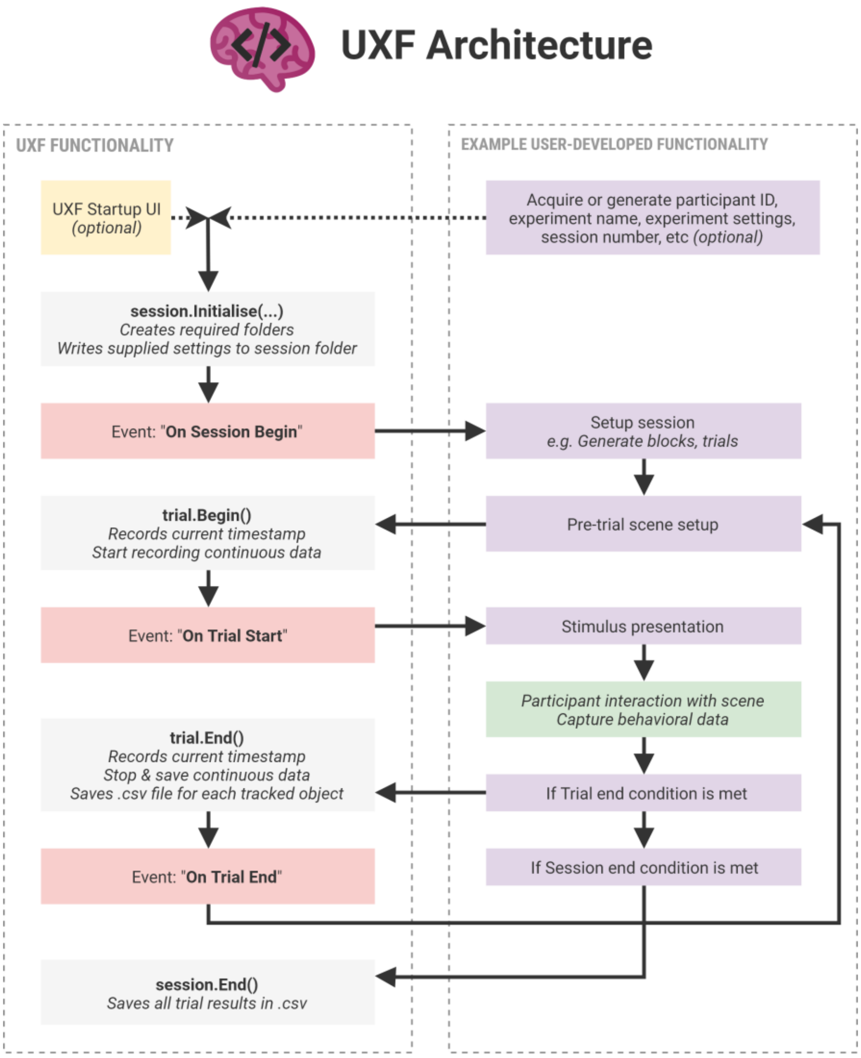

UXF Architecture
UXF FUNCTIONALITY
EXAMPLE USER-DEVELOPED FUNCTIONALITY
UXF Startup UI (optional)
Acquire or generate participant ID, experiment name, experiment settings, session number, etc (optional)
session.Initialise(...) Creates required folders Writes supplied settings to session folder
Event: "On Session Begin"
Setup session e.g. Generate blocks, trials
trial.Begin() Records current timestamp Start recording continuous data
Pre-trial scene setup
Event: "On Trial Start"
Stimulus presentation
Participant interaction with scene Capture behavioral data
trial.End() Records current timestamp Stop & save continuous data Saves .csv file for each tracked object
If Trial end condition is met
Event: "On Trial End"
If Session end condition is met
session.End() Saves all trial results in .csv

Figure 4.4 – Structure of a typical task developed with UXF. The left panel shows functionality present in UXF, with functionality a researcher is expected to implement shown on the right panel. The framework features several “events” (shown in red) which are invoked at different stages during the experiment; these allow developers to easily add behaviours that occur at specific times, for example presenting a stimulus at the start of a trial.

4.3.6. Multithreading file I/O

Continuous measurement of variables requires large amounts of data to be collected over the course of the experiment. When using a VR head-mounted display, it is essential to maintain a high frame rate and keep stutters to a minimum to minimise the risk of inducing sickness or discomfort on the participant. Handling of tasks such as reading and writing to file may take several milliseconds or more depending on operating system background work. Constant data collection (particularly when tracking the movement of many objects in the scene) and writing these data to file therefore poses a risk of dropping the frame rate below acceptable levels. The solution is to create a multi-threaded application which allows the virtual environment to continue to be updated whilst data are being written to files simultaneously in a separate thread. Designing a stable multithreaded application imparts additional technical requirements on the researcher. UXF abstracts file I/O away from the developer, performing these tasks automatically, with a multithreaded architecture working behind the scenes. Additionally, the architecture contains a queueing system, where UXF queues up all data tasks and writes the files one-by-one, even halting the closing of the program to finish emptying the queue if necessary.

4.3.7. Cloud-based experiments

UXF is a standalone, generic project, and as such, it does not put any large design constraints on developers using it. This means that UXF does not have to be used in a traditional lab-based setting, with researchers interacting directly with participants; it can be used for data collection opportunities outside of the lab, by embedding experiments within games or apps that a user can partake in at their discretion. Data are then sent to a web server where it can later be downloaded and analysed by researchers (Figure 4.5). Recently these cloud-based experiments have become a viable method of performing experiments on a large scale.

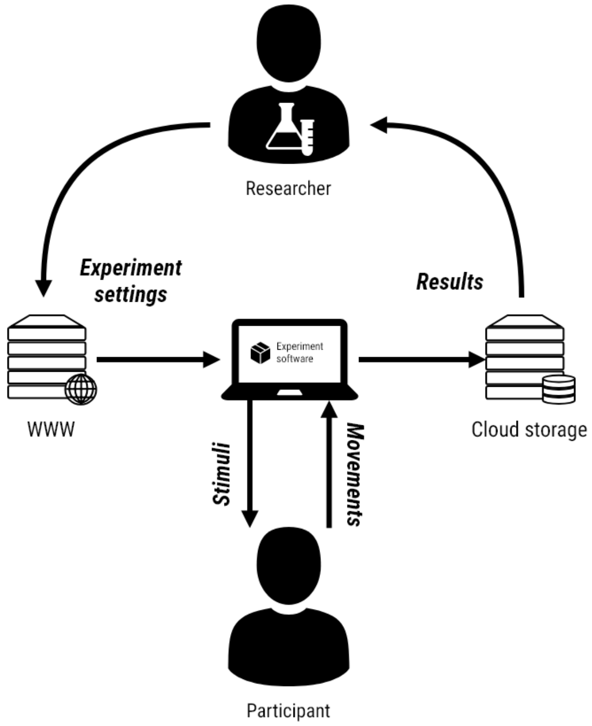


Figure 4.5 – Experiment in the cloud. A piece of software developed with UXF can be deployed to an internet connected device. Researchers can modify experiment settings to test different experimental manipulations over time, which are downloaded from the web by the client device upon running a UXF experiment. As the participant partakes in the experiment, they are presented with stimuli, and their movements are recorded in the form of behaviours/responses or continuous measurement of parameters like hand position. Their results are automatically and securely streamed up to a server on the internet, of which the researcher can periodically retrieve data from.

UXF can be used in cloud-based experiments (Figure 4.5) using two independent pieces of software that accompany UXF:

1. *UXF S3 Uploader* allows all files that are saved by UXF (behavioural data, continuous data, logs) to be additionally uploaded to a location in Amazon’s Simple Storage Service as setup by a researcher. This utilizes existing UXF functionally of setting up actions for after a file has been written; and so a developer could potentially implement uploading the files to any other storage service.

2. *UXF Web Settings* replaces the default UXF functionality of selection of experiment settings via a user interface, to the settings being accessed automatically from a web URL by the software itself. This allows a deployed experiment (e.g. via an app store, or simply transferring an executable file), to be remotely altered by the researcher, without any modification to the source code. Settings files are stored in json format and would usually be of a very small file size so can be hosted online cheaply and easily.

A developer can implement neither, either, or both, depending on the needs of the research. For lab-based experiments, neither are required. For experiments without any need to modify settings afterwards, but with the requirement of securely backing up data in the cloud, (1) can be used. If a researcher wants to remotely modify settings but has physical access to the devices to retrieve data, (2) can be used. For a fully cloud-based experiment without direct researcher contact with the participant both (1) and (2) can be used. This has been successfully tried and tested in the context of a museum exhibition, where visitors could take part in VR experiments, with the recorded data being uploaded to the internet. Both UXF S3 Uploader and UXF Web Settings are available as open source Unity packages.

4.3.8. Case study

One classic question in human behavioural research has related to the information used by adults and children when maintaining posture (Thomas & Whitney 1959; Edwards 1946). To investigate the contribution of kinaesthetic and vision information when both are available, four decades ago Lee and Aronson (1975) used a physical ‘swinging’ room to perturb the visual information provided by the walls and ceiling whilst leaving the kinaesthetic information unaffected (only the walls and ceiling swung, and the floor did not move). This experiment demonstrated the influence of vision on posture, but the scale of the apparatus meant that it could only ever be implemented in a laboratory setting. The approach was also subject to measurement errors and researcher bias (Wann, Mon-Williams & Rushton 1998). More recently, conventional computer displays have been used to explore the impact of vision on posture (e.g. Villard et al 2008) and this method has addressed issues of measurement error and researcher bias but still remains confined to the laboratory.

The ability to create a virtual swinging room in a VR environment provides a test case for the use of UXF in supporting behavioural research and provides a proof-of-concept demonstration of how large laboratory experiments can be placed within a non-laboratory setting. Here, we used the head tracking function as a proxy measure of postural stability (as decreased stability would be associated with more head sway; Flatters et al., 2014). In order to test the UXF software, we constructed a simple experiment with a within-participant component (whether the virtual room was stationary or oscillating) and a between-participant factor (adults vs children). We then deployed the experiment in a museum with a trained demonstrator and remotely collected data on one hundred participants.

The task was developed in the Unity game engine with UXF handling several aspects of the experiment including; Participant information collection, Settings, Behavioural data and Continuous data. *Participant information collection*: The UXF built-in user interface was used to collect a unique participant ID as well as the participant's age and gender. This information was stored in a CSV participant list file. This list was subsequently updated with participant height and arm-span as they were collected in the task. *Settings*: A settings file accompanied the task that allowed modification of the assessment duration as well as the oscillation amplitude and period without modifying the code. Settings for each trial were used to construct the environment to facilitate the requested trial condition. *Behavioural data*: While there were no dependant variables that were directly measured on each trial, the UXF behavioural data collection system output a list of all trials that were run in that session, as well as the vision condition for that trial. *Continuous data*: UXF was configured to automatically log the HMD position over time within each trial, which was then used offline for the stability measure calculation. UXF split the files with one file per trial which was designed to make it easy to match each file with the trial condition the file was collected under.

Methods

Fifty children (all under <16 years of age; mean age: 9.6 years; SD: 2.0 years) and 50 adults (mean age: 27.5 years; SD: 13.2 years) took part in the study. Participants were recruited from either the University of Leeds participant pool (adults) or were attendees at the Eureka! Science museum (children and adults) and provided full consent. A gaming-grade laptop (Intel Core i5-7300HQ, Nvidia GTX 1060) in addition to a VR HMD (Oculus Rift CV1) and the SteamVR API, a freely available package independent of UXF (Valve Corporation, 2018) were used to present stimuli and collect data. The HMD was first calibrated using the built-in procedure, which set the virtual floor level to match the physical floor.

After explaining task requirements, the demonstrator put the HMD on the participant’s head (over spectacles if necessary) and adjusted it until the participant reported it was comfortable and they could see clearly. Participants were then placed in the centre of a simple virtual room (height: 3m, width: 6m, depth: 6m) with textured walls and floors (Figure 4.6). Height was measured as vertical distance from the floor to the “centre eye” of the participant (as reported by the SteamVR API) and this value was used to place a fixation cross on the wall at the participant’s height.

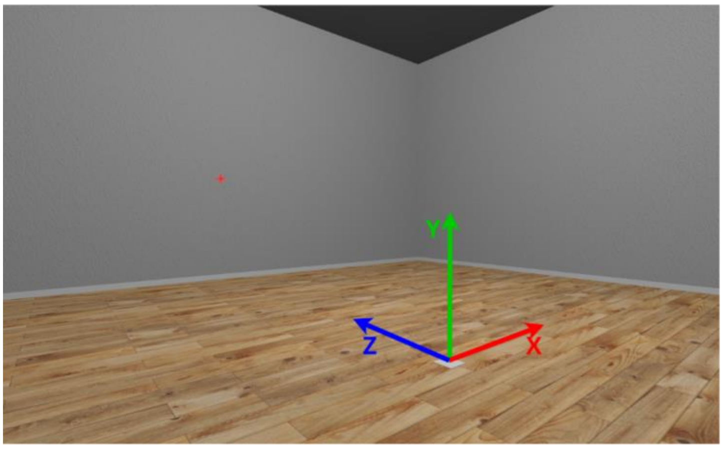


Figure 4.6 – Screenshot from inside the virtual room. Arrows indicate the three axes as well as the origin. The red fixation cross is shown on the wall.

The task comprised two 10 second trials performed in a random order. The *normal* condition asked participants to stand still and look at a fixation cross placed on the wall. In the *oscillating* condition, the participants were given the same instructions, but the virtual room oscillated in a sinusoidal fashion (rotating around the x axis) with an amplitude of 5° and a frequency of 0.25Hz. The oscillation was performed about the point on the floor at the centre of the room, in effect keeping the participant's feet fixed in-place. Participants were not explicitly informed about the room oscillation. The position of the HMD inside the virtual room was logged at a rate of 90Hz during each of the two trials. The path-length of the head was used as a proxy measure of postural stability (sum of all point-to-point distances over a trial).

4.3.9. Results

No participants reported any feelings of sickness or discomfort during or after taking part in the task. A mixed-model design ANOVA (2 [Age: Adult vs Children] x 2 Vision Condition [Normal vs. Oscillating]) found no interaction, F(2, 98) = 0.34, p = .562, η^2_G = .001, but revealed main effects of Vision, F(2, 98) = 7.35, p = .008, η^2_G = .016 and Age, F(1, 98) = 9.26, p = .003, η^2_G = .068, thus replicating previous work on the contribution of visual information on postural stability (Flatters et al., 2014).

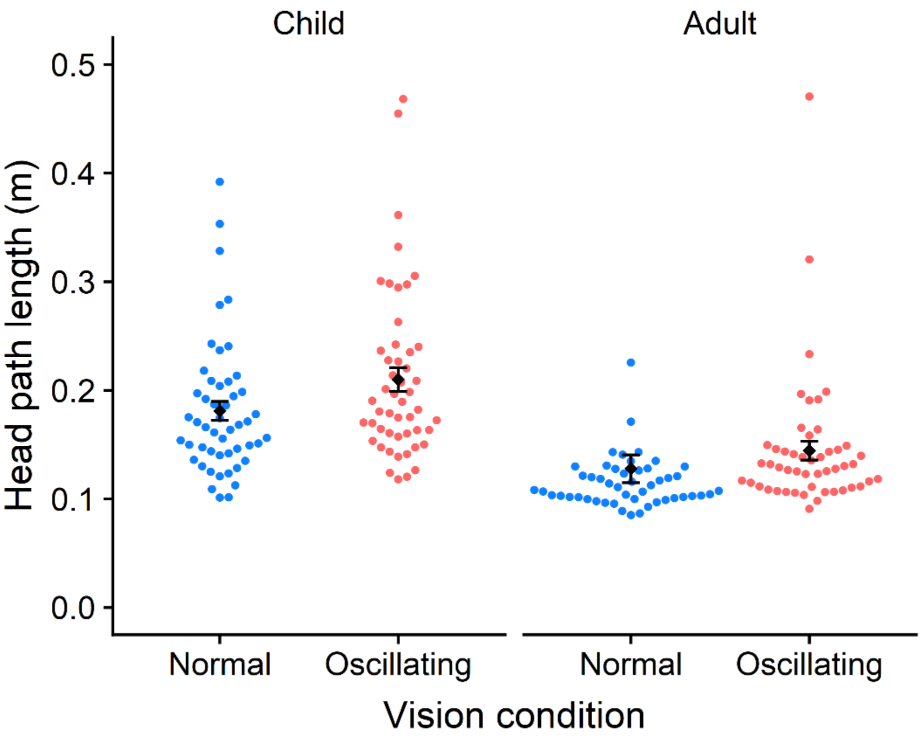


Figure 4.7 – Head path length (higher values indicating worse postural stability) as a function of vision condition. The two conditions were ‘normal’ (static virtual room) and ‘oscillating’ (oscillating virtual room). Postural stability was indexed by the path length of head movement in meters (measured over a 10 second period). Adults showed significantly different path length overall compared to children (shorter – indicating greater stability). Error bars represent +/- 1 SEM.

4.3.10. Summary

We have created an open source resource that enables researchers to use the powerful games engine of Unity when designing experiments. We tested the usefulness of UXF by designing an experiment that could be deployed within a museum setting. We found that UXF simplified the development of the experiment and produced measures in the form of data files that were in a format that made subsequent data analysis straight forward. The data collected were consistent with the equivalent laboratory-based measures (reported over many decades of research) whereby children showed less postural stability than adults, and where both adults and children showed greater sway when the visual information was perturbed. There are likely to be differences in the postural responses of both adults and children within a virtual environment relative to a laboratory setting and we would not suggest that the data are quantitatively similar between these settings. Nonetheless, these data do show that remotely deployed VR systems can capture age differences and detect the outcomes of an experimental manipulation.

Our planned work includes maintaining the software for compatibility with future versions of Unity, and refactoring UXF so that it works on a wider range of platforms (e.g. mobile devices, web browsers, augmented reality devices, standalone VR headsets). Features may be added or modified if a clear need for a feature arises. The project is open source, thus allowing researchers in the field to implement and share such additions.

4.4. Availability

UXF is freely available to download via GitHub as a Unity Package (github.com/immersivecognition/unity-experiment-framework), and currently can be integrated into Unity tasks built for Windows PCs. Documentation and support is available on the GitHub wiki (github.com/immersivecognition/unity-experiment-framework/wiki). The package is open sourced under the MIT licence. Related packages UXF S3 Uploader and UXF Web Settings are available via the same GitHub link.

5. Motor Bliss and Visual Error Amplification

5.1. Overview

In Chapters 2 and 3 we examined motor learning by exposing participants to a novel forcefield and applying interventions that increase workspace exposure through manipulating task space variability and force-induced positional error. In Chapter 4 we reported the development of a new software framework that allows researchers to readily develop experiments to study motor learning in virtual environments. In this chapter, we take advantage of this new software to create a complex motor control task and tackle two related research questions that have emerged from the experimental work reported thus far. First, we delve deeper into the role of the role of variability on motor learning by focussing on solution space variability (c.f. Chapter 3). Second, employing the same task, we develop a novel visual error amplification intervention to investigate whether the augmentation of this signal can artificially accelerate learning without the use of haptics.

5.2. Introduction

The sensorimotor system is remarkable for numerous reasons, but perhaps none more so than the fact that for any desired action, there are often multiple to infinite ways of achieving the goal. One of the earliest and certainly most influential formalisations of this observation comes from Nikolai Bernshteĭn (1967), whose examination of the kinematics of

Blacksmiths led to the phrase “repetition without repetition”- even when the high-level goal remains constant, redundancy in elemental variables produces differences in action execution. This has variously been described as the “degrees of freedom problem” or the “problem of motor redundancy”. The variability that arises from this redundancy appears to be an intrinsic characteristic of human sensorimotor control, but more recent interpretations have reversed course on the idea that this is in any way *problematic* per se. Variability, it seems, is not a bug in the system, but a feature (Tumer and Brainard, 2007; Huang and Shadmehr, 2009; Wu et al., 2014). This “bliss of motor abundance” provides a balance between stability and flexibility (Latash, 2000) required to navigate through the world around us. This reformulation has been coupled with a surge in research interest on the utility of motor variability.

Bernshteĭn’s early description was predicated on the hypothesis that individual muscles are not controlled in isolation, but that actions are planned and executed in terms of higher-level movements. At the outset of learning a new motor task, we do not have the high-level control strategies that would allow us to work with the redundancy. Instead, it seems we employ strategies to reduce the degrees of freedom by either combining degrees of freedom (Li, 2006) or eliminating them completely (Newell, 1991).

The Uncontrolled Manifold Hypothesis (Scholz and Schöner, 1999) describes these different types of variability in terms of the constituent degrees of freedom used to perform the task. Task space variability concerns differences in position that do affect performance in the task at hand. For example, variability in the position (between attempts) of a thrown dart on a dartboard, assuming the player is attempting to hit a particular point on the board, is task-space variability.

Evidence from Wu et al. (2014), suggested that participants with high variability in the task space (affecting performance) in the early stages of learning improve their performance through training when compared to low variability participants. However, the variability observed by Wu & colleagues is only of the type that affects performance in the task, and recent analyses found mixed results on the impact of variability in motor learning (He et al., 2016). Cardis, Casadio, & Ranganathan (2017) facilitated variability both in the task space (affecting performance) and in the redundant parameter space or null space (not affecting performance), to examine if either of these forms of variability could enhance learning. They found that any variability intervention was detrimental to learning, even though participants subjected to variability were exposed to more potential solutions.

Specific combinations of joint angle movements can produce no movement in the task space (i.e. movement that would help or hinder us achieving a goal such a reaching for a target). The subspace, or manifold, that these degrees of freedom combinations lie on is sometimes called the null-space, or the uncontrolled manifold.

For a given task, a solution used could lie anywhere in the null-space. For example, there are an infinite number of elbow orientations that would all allow for reaching a target perfectly accurately due to the redundancy in the joints of the human arm. The space in which these solutions lie in terms of joints is the null space. Thus, when considering the changes in this solution between trials, a measure of variability can be made indicating how consistent (or not) these solutions are across trials. A high null-space variability and low task-space variability is a so-called “synergy”, i.e. a grouping of (e.g.) joints that work in-tandem to reduce task-space variability, and a marker of high skill (Latash and Anson, 2006).

It remains unclear if the natural variability that is present in the early stages of learning is a by-product of poor performance, or if this variability is also a somewhat deliberate strategy by the motor system to find new solutions (i.e. “motor babbling”). If it is the latter, an intervention imposing artificial variability such as those tested by Cardis et al. (2017) may not lead to the same retention of those discovered solutions. The motor system may impose a more optimal variability pattern that explores the solution space that benefits the learner, compared to an artificially imposed variability. Therefore, we set out to investigate individuals’ natural null-space variability, how it relates to individuals’ learning, and how it changes over time.

5.2.1. Manipulating errors

The learning of a motor task can, as we identified in a previous chapter of this thesis, also be accelerated artificially through the presentation of disruptive forces (Laura Marchal-Crespo et al., 2014). In previous work, we raised the possibility that learning in these scenarios may be a by-product of the increased task-related information experienced by the participant, as disruptive forces facilitate exploration of the domain of the task. An experiment directly manipulating the information available to the participant, without manipulating error, found this type of information exposure had no impact on learning.

A second hypothesis emerging from this work proposed that the amplified error observed through presentation of disruptive forces enhances the error signal, thus driving learning. This enhanced signal could be observed through means of an increased signal-to-noise ratio in our perception of the error vector, or increased attentional resources directed towards correcting the error when it is seen to be large and if so, should operate independent of the haptic disruptive forces employed in previous experiments. Thus, we set out to increase error by visually amplifying the error signal and examining its impact on learning, and specifically how these impact on our selection of execution parameters.

5.2.2. A Novel Motor Learning Task

Studying trends in changes in null-space variability are generally difficult in most motor control experiments due to the complexity of measuring movements in terms of their constituent joint angles. To this end, we created a novel motor task that is both complex enough that changes in performance (learning) can be easily observed, but also simple enough that the fundamentals of solving the degrees of freedom problem can be examined (thus allowing us to capture changes in variability in the null-space). We presented participants with a task that involved a novel mapping of 4 hand rotations to the movement of a 2D cursor.

Our primary goal was to explore whether null-space variability (c.f. task-space variability; Wu et al., 2014) could predict future learning. In addition to this, in line with Bernshteĭn's ideas on reducing the degrees of freedom problem, we predicted that those who learn to resolve the degrees-of-freedom problem through reducing the solution space (e.g. by holding one axis constant) would be the fastest learners. Finally, predicated on our previously reported experiments, we hypothesized that performance after training under an enhanced error condition would be improved (when the manipulation is removed). We had no a priori predictions about how this manipulation may interact with variability but expected these two independent manipulations to contribute to the learning of our novel motor learning task.

5.3. Methods

5.3.1. Participants

18 participants (10 female, 8 male, all right handed) were recruited from the University of Leeds School of Psychology Participant Pool. The School of Psychology Ethics Board at the University of Leeds approved the research.

5.3.2. Procedure

We examined motor learning through a virtual reality environment in which participants needed to successfully resolve a novel mapping between hand and cursor as quickly as possible. To this end, participants were invited to the Experimental Psychology Research Unit Laboratory where they were seated on an armed chair and wore an Oculus Rift virtual reality head-mounted-display (VR HMD) and had an Oculus Touch VR controller in each hand. Care was taken to ensure the HMD was correctly mounted for comfortable viewing. The virtual environment was set up such that the origin was positioned above the chair with the forward direction (+z) aligned with the forward direction of the chair. The task was developed with the Unity game engine version 2017.3 (Unity Technologies, 2018) and the Unity Experiment Framework (Brookes et al., 2019).

After the task procedure was explained to the participant, the session began with participants rotating their hands to the default orientation (0, 0, 0 degrees) with a tolerance of +/-3 degrees in all axes, visually guided by aligning a solid cuboid with a wireframe cuboid. The shape turned green, the workspace turned amber, and a small haptic pulse was emitted when the hand was in the correct orientation. After holding both hands in the default orientation for duration of 0.75 - 1.25 seconds (randomly sampled), a whistle sounded, the workspace turned green, and the trial began. A trial entailed moving a cursor (blue sphere) towards a target (red sphere; both of 3cm diameter) which appears in one of 6 locations in the workspace. The workspace was in the x-y plane 0.5m away from the origin, and entirely in the participant's field of view without turning their head. The cursor was moved by rotating the hands, and according to the novel mapping, the cursor would move around the workspace. The mapping is a function of the x and y rotations of the hands (pitch and yaw respectively), with the roll (and position in space) of the hands not contributing to the cursor's movement. The participant must hold the cursor on the target for 0.5 seconds (max centre to centre distance = 3cm; progress shown as a horizontal green bar above the target) before completing the trial. Upon completing the trial, the target and cursor disappeared and once again the participant must align their hands to the default orientation using the cuboid guides. This effectively sets the cursor position at the midpoint of the workspace ready for the next trial. The participant was told they will be assessed on "the time it takes to reach to the target as well as the smoothness and accuracy of their movements". There were additional interventions that could occur during a trial; first, if the participant were to rotate their hands to an orientation outside the required range ($-50 \leq x \leq 50$, $-100 \leq y \leq 100$, $-30 \leq z \leq 30$), the cursor and target disappeared, and a warning screen shows until they return them to within the acceptable range. This kept the hands in a comfortable range and avoided an issue where the cursor would wrap around the workspace if participants were to rotate beyond 180 degrees. Secondly, an assistance screen was presented below the workspace if participants could not complete a trial within 20 seconds. This showed an image of the two hands and arrows depicting the required rotation axes. Timings, angle limits & distances were initially selected by examining similar methods in the publicly available scientific literature. Through pilot testing, these parameters were refined to ensure the task was comfortable for all users.

The experiment comprised 10 sessions (one per weekday for 2 weeks, always starting on a Monday). Within each session, there were 4 blocks each with 36 trials (Figure 5.1A); each session lasted around 20 minutes. The 1st block on each Monday, and the 4th block on each Friday were classified as “assessment” blocks. All other blocks were classified as “training” blocks.

Of the 36 trials in each block, there were 12 of repetitions of each of the three targets for that block, randomly shuffled within the block (but maintaining the same order for between participants). One of two target sets, normal (targets A, B & C) and alternative (targets D, E, F), were used in blocks classified as training and assessment respectively (Figure 5.1C). This two different target sets ensured that any learning measures were measures of performance improvement in the ability to manipulate the cursor through the novel mapping, rather than memorisation of poses required to move towards the targets they trained with. There was no indication to the participant of transition between blocks, the participant experiences a continuous series of trials.

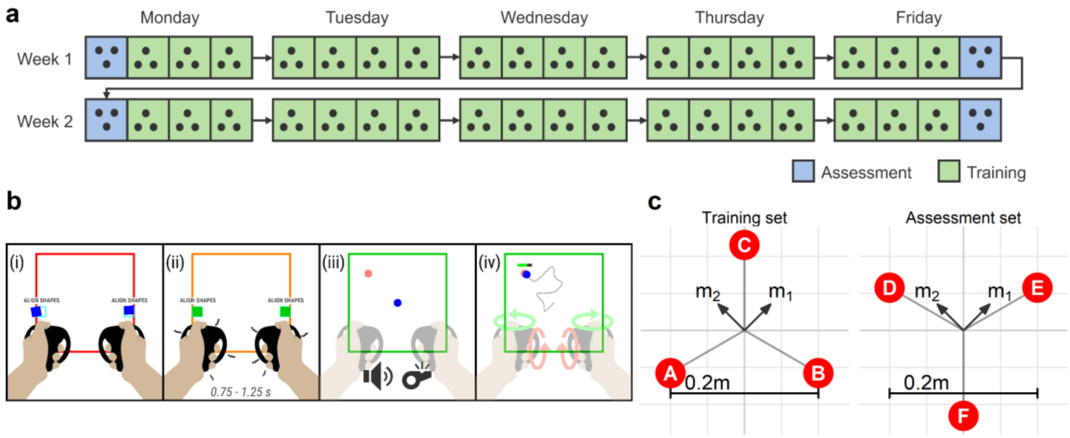


Figure 5.1 – Experimental procedure. (a) The experiment contained 10 sessions, one per weekday for 2 weeks. Each day consisted of 4 blocks (except for Week 2 Friday). Monday Block 1 and Friday Block 4 were assessment blocks. Each block contained 36 trials (with 12 of each of the 3 targets for the given target set). (a) Trial procedure. [i] The participant aligned their hands to the default orientation guided by a cuboid above each hand. [ii] When aligned, haptic feedback was felt in each hand, and the participant must hold the position for a random period. [iii] The blue cursor and red target stimuli appeared along with a whistle sounding, and the hand controllers are made invisible. [iv] The participant must rotate their hands to move the cursor towards the target and stay there until the progress bar fills (dotted trail for illustrative purposes only). The process then repeats from the first panel, with a randomly selected target location. (c) There are 6 possible target locations on the workspace, split into 2 target sets. The Assessment Set is used in Assessment blocks, and the Training Set used in Training blocks.

5.3.3. Novel mapping

The cursor was manipulated through the rotating a controller in each hand (each controller contains a high precision inertial measurement unit; IMU). The rotations of these controllers were then converted into an x, y coordinate of the cursor in the workspace according to the novel mapping. The inputs to novel mapping were the x & y components of rotation in each hand ($r_{Lx}, r_{Ly}, r_{Rx}, r_{Ry}$). These were converted into a position in “task space” m_1 and m_2, by multiplying by scaling coefficients (1/175 for x rotation, 1/350 for y rotation) dictating the meters the cursor should move per degree of rotation

$$\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{175}(r_{Lx} + \ r_{Rx}) \\ \frac{1}{350}\left(r_{Ly} + \ r_{Ry}\right) \end{bmatrix} \qquad 5$$

This essentially made the task a search through 4-dimensional space to find a subspace which causes the cursor to meet the target. Crucially, this subspace is a 2-dimensional plane creating redundancy in the task. i.e., there exist a range of parameter combinations which cause the cursor to meet the target. Note that this subspace or plane's location in the parameter space was defined by the target location. Also note that the y rotation of the left hand (r_{Ly}) was inverted in direction, to make the rotation contribution between hands symmetrical (Figure 5.2b).

To convert to the cursor position in the workspace ($\mathbf{c}$), rotate position in movement axes $\mathbf{m}$ by 45° was rotated, so that x and y rotations do not move exactly in the x and y directions respectively, but along axes offset 45° from these. This makes the mapping rules more difficult to solve and restricts immediate development of high-level strategies for the task.

$$\mathbf{c} = R(45°)\ \mathbf{m} \qquad 6$$

$R(\theta)$ is the transformation matrix which rotates a vector of points by angle θ about the origin.

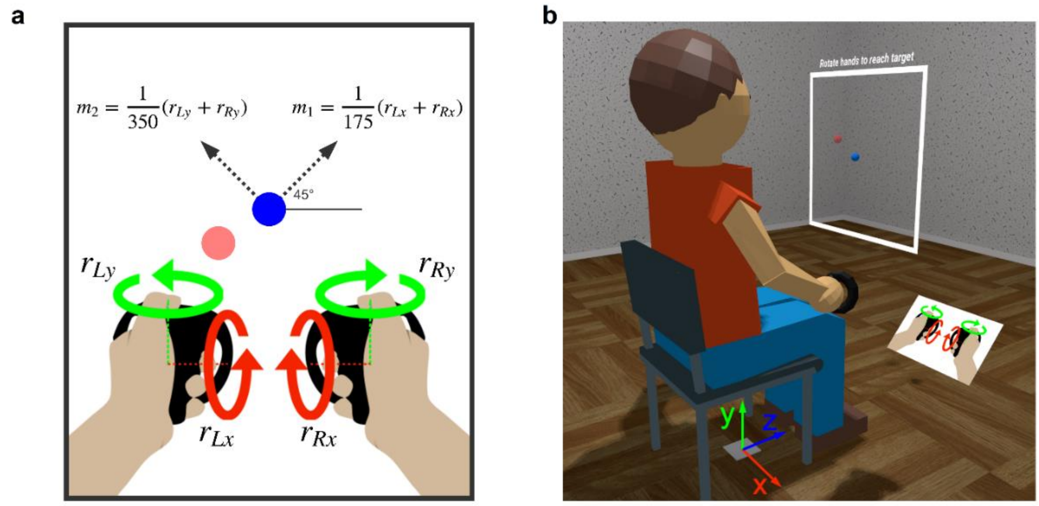

a
b
$m_2 = \frac{1}{350}(r_{Ly} + r_{Ry})$
$m_1 = \frac{1}{175}(r_{Lx} + r_{Rx})$
45°
r_{Ly}
r_{Ry}
r_{Lx}
r_{Rx}
Rotate hands to reach target
y
z
x

Figure 5.2 – Experiment setup in virtual reality. (a) The participant used a VR input device in each hand, rotations of which mapped on to axes m_1 and m_2 to control a cursor. The 4 rotations that were used are shown, including r_{Ly} which is inverted to provide bilateral symmetry (b) The participant was seated wearing a head-mounted display and placed in a virtual empty room and were seated above the origin in a left-handed coordinate system. Target and cursor stimuli were spheres of diameter 3cm and were presented on a plane parallel to X-Y at z = 0.5m.

Fulfilling the demands of the task (i.e. m_1 and m_2 produced via hand movements must cause the cursor to meet the target) can be completed though any number of possible solutions due to the redundancy in the mapping. The chosen solution can be quantified by calculating the position in axis orthogonal to the task space axis (either m_1 or m_2). We define these parameters associated with the two task space axes as p_1 and p_2 respectively,

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{175} r_{Lx} - \frac{1}{350} r_{Ry} \\ \frac{1}{350} r_{Ly} - \frac{1}{175} r_{Rx} \end{bmatrix}. \qquad 7$$

This means that $\mathbf{p}$ represented our null-space and $\mathbf{m}$ is our task-space, and any given value of $\mathbf{m}$ (e.g. a demand target position) we can express chosen solution as this two-element vector $\mathbf{p}$.

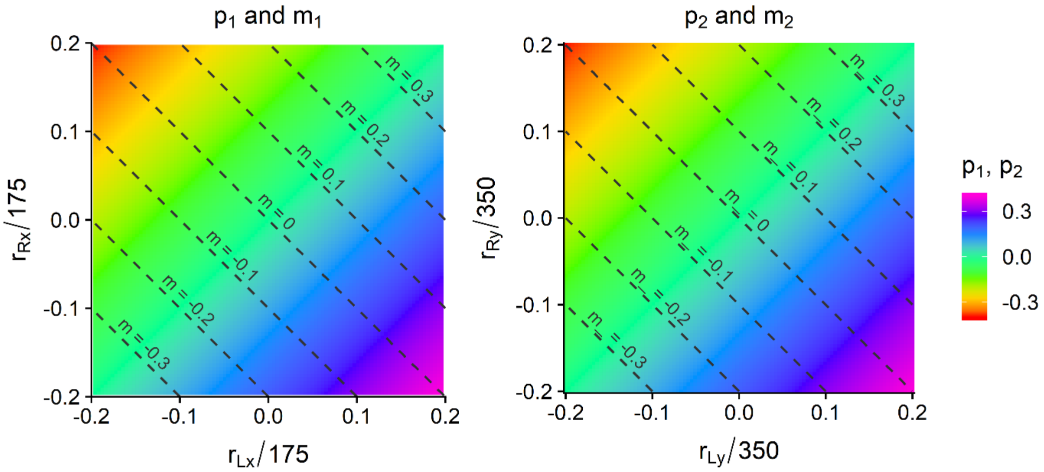


Figure 5.3 – Graphical examples of the contribution of control parameters to task-space parameters (m_1 and m_2) and null-space parameter (p_1 and p_2). Here, dotted lines show contours of values for task-space parameters (i.e. a "physical" position of the cursor). By requirements of completing the task, the participant must, via control of the 4 control parameters (rotations), move between targets that are situated in task-space. The colouring represents the values for null-space parameters, or a quantification of the "pose" (i.e. control parameter combinations which do not affect physical cursor position). A change in only a null-space parameter would result in a line with gradient -1 here (e.g. the dashed lines), where a change in a control parameter is compensated for with the equal and opposite change in the other constituent parameter, hence leaving the associated task-space parameter unchanged.

5.3.4. Error amplification

During training blocks, the cursor position was manipulated for participants in the "error amplification" group. This shows the cursor to be further from the target than it should be (but always maintaining the same direction). In these trials instead a "fake" cursor $\mathbf{c_f}$ will be shown at an offset from the real cursor position $\mathbf{c}$. The 2nd derivative of the sigmoid function S is used to generate the offset between $\mathbf{c}$ and $\mathbf{c_f}$:

$$S''(x) = \frac{e^x(e^x - 1)}{(1 + e^x)^3} \quad 8$$

$S''(x)$ is used in combination with parameters r (range) and A (amplitude), which were set, in this task, to 0.04 and 0.7 respectively. The fake cursor position $\mathbf{c_f}$ is calculated by taking the magnitude and direction of the real error $\mathbf{e}$ (difference between cursor and target positions) and manipulating it to create a fake error $\mathbf{e_f}$. This fake error vector $\mathbf{e_f}$ which is then added to the target position $\mathbf{t}$ to generate the fake cursor position $\mathbf{c_f}$

$$\mathbf{e_r} = \mathbf{c_r} - \mathbf{t} \quad 9$$

$$\mathbf{e_f} = A \cdot S''\left(\frac{|\mathbf{e_r}|}{r}\right) \cdot \hat{\mathbf{e}}_\mathbf{r} \quad 10$$

$\hat{\mathbf{e}}_\mathbf{r}$ is the normalized vector $\mathbf{e_r}$.

$$\mathbf{c_f} = \mathbf{t} + \mathbf{e_f} \quad 11$$

The use of this method for manipulating error ensures the fake cursor moves smoothly and allows the user to maintain a sense of control over the cursor. Importantly, the real cursor position is always used as the marker to trigger the end of a trial on successful reaching to and holding on a target. i.e., the real error $\mathbf{e_r}$ must be below 3cm for a contiguous 0.5s for the trial to end regardless of the fake cursor position. The use of the real error over the fake error ensures the task does not get more difficult with error amplification due to the mechanics of the task; it is only visual information that is altered between groups.

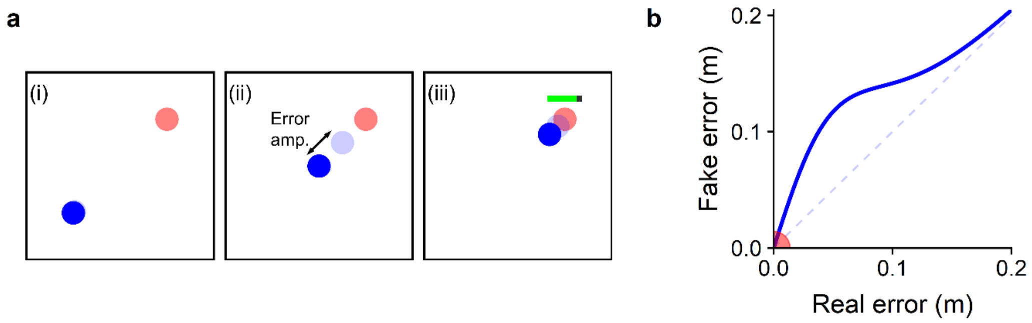


Figure 5.4 – Error amplification. (a) Illustration of a single error amplified trial. [i] A red target sphere appears at a position on the workspace, while the participant controls the blue target using the novel rotation mapping. [ii] In the error amplification group, error is amplified by means of showing the fake cursor (shown here as blue) and hiding the real cursor (shown here as pale blue) during training. [iii] Upon moving the cursor to the target, a progress bar begins to fill up; after 0.5s the target then moves to a different location to begin the next trial. (b) Fake cursor has an amplified error as a function of the real cursor error during the training in the error amplification condition. With parameters r and A set to 0.04 and 0.7 respectively, fake cursor position is seen here to be amplified the most around an error of ~7cm from the target.

5.3.5. Metrics

To assess motor performance, we used movement time for each trial. Performance improvement was calculated by subtracting the mean post-training score from the mean pre- training score (participant means of respective block). Exponential learning curves were fit to the performance scores in the training blocks using the equation

$$\text{mean movement time} = \ ae^{b \times \text{block number}} + c, \qquad 12$$

including only training blocks, excluding assessment blocks.

To assess exploration of various solutions to the reaching of a target, first the solution to each target (the orientation of the hands when at the endpoint of the trial) was quantified. There were 4 control parameters the participant has access to $(r_{Ly}, r_{Lz}, r_{Ry} \,\&\, r_{Rz})$ which contribute to two task-space parameters $(m_1 \,\&\, m_2)$.

The two task-space parameters are each a function of only two of the control parameters ($m_1 = f(r_{Lx}, r_{Rx}),\ m_2 = f(r_{Ly}, r_{Ry})$). Since the value of each of the task-space parameters are each a sum of the (distance) contribution each of its constituent control parameters, we can create a unique meta-parameter for each task-space parameter. These meta-parameters represent the location of this solution in a “null-space”, i.e. the axis (perpendicular to the task space) within the control parameter space which would incur no change in task-space parameters if traversed along. The two-to-one mapping of control parameters to task-space parameters results in two distinct null-space parameters. Effectively, this task consists of two independent, redundant systems, performed simultaneously. Trials in this task can be thought of as searching for the two correct values of the task space parameters via control of the four control parameters. With reference to Equation 5, we can calculate the null-space parameters by subtracting one of the constituent control parameter contribution from the other, i.e.

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{175}(r_{Lx} - r_{Rx}) \\ \frac{1}{350}(r_{Ly} - r_{Ry}) \end{bmatrix} \qquad 13$$

Null-space parameters p_1 and p_2 can, by definition, be modified independently from task-space parameters m_1 and m_2. They can be thought of as describing the “pose” of the hands for the associated task-space position. Task- and null-space parameters can be represented graphically when plotting possible values of the constituent control parameters against each other (Figure 5.3).

The value of the null space parameters can be captured at each point in time with Equation 11. We can examine values of null-space parameters at the endpoint of each trial (i.e. when a solution is found). Here, the task space parameters are equal to the target position (within the allowable error radius). Between-trial null space variability of the movement endpoints was quantified by calculating the distance to the mean parameter position for the given target for each block (therefore the variability across 12 trials),

$$\text{variability} = \sqrt{\frac{1}{1-N}\sum_{i=1}^{N}\left|\begin{pmatrix}p_{1_i}\\ p_{2_i}\end{pmatrix}-\begin{pmatrix}\bar{p}_1\\ \bar{p}_2\end{pmatrix}\right|^2} \qquad 14$$

Equation 14 essentially calculates standard deviation but uses the 2-dimensional distance to the mean rather than the 1-dimensional distance. Then, these are averaged across the 3 targets to give an overall variability per block. Other measures were considered but ultimately had issues, such as standard deviation in each direction (results in directional artefacts) and area of a best-fit ellipsis shape (variability in only a single direction would result in an area of 0).

These data can then be used in the same way as the performance metric: reduction in exploration in post-training subtracted from pre-training, and rate of change of exploration by fitting values to an exponential curve in the form shown in Equation 12. One participant was excluded from this curve fitting analysis after the exponential fitting failed to converge (inspection revealed abnormally high variability in a single block during training).

5.4. Statistical analyses

5.4.1. Error amplification

To examine if there was an overall effect of the error amplification intervention on learning, an ANCOVA (Error Amp vs Control) for the final assessment movement time (with pre-training assessment movement time as a covariate), was performed.

Independent t-tests on the parameters of the curve fit between the two conditions, were conducted to examine if the performance during training was different in terms of a (initial performance relative to floor), b (indicator of learning rate) and c (floor level), from Equation 12. Effect sizes (*Cohen's d*) are reported where appropriate. All data met assumptions of normality through assessment by histogram, Q-Q plots, and Shapiro-Wilk tests.

5.4.2. Parameter space selection

To examine whether variability within participants changed over time, exponential curves were fit to the variability measure within training blocks (same process as performance curves, above). Specifically, a one-sample t-test on the b value of the exponential fits was used, which would have a value of 0 if there was no change over time (negative for decrease in variability, positive for increase in variability).

Spearman’s rho was calculated to examine the relationships between measures gathered for each participant. Spearman’s was selected over Pearson’s because (a) were interested in monotonic relationships; and (b) these data did not meet assumptions of normality, with outlier datapoints making identifying linear relationships difficult. Specifically, 4 correlations were of interest:

- Initial endpoint null-space variability vs initial performance, to validate if variability and performance are independent measures of behaviour. This is to ensure there are no confounds arising from the possibility that (for example) low variability is an inherent feature of good performance. Note this analysis is performed on the participant values for the initial assessment, which uses the assessment target set.

- Final endpoint null-space variability vs final performance, to test two different perhaps conflicting observations in motor learning. Here, we asked whether the participants who managed to reduce their variability the most end up performing best? Or were the best performers able to utilise their understanding of the task to produce high variability, with no cost to performance (a “synergy”)? Note this analysis is performed on the participant values for the final assessment, which uses the assessment target set.

- Initial endpoint null-space variability vs learning rate. This tests a claim similar to that demonstrated by Wu et al. (2014), in that the initial variability can predict subsequent learning rate. Here, we explore whether null space variability, which can exist independent of performance, bears a relationship with later performance.

- Variability reduction rate vs learning rate, to assess whether those who were the fastest at reducing variability are also the fastest at reducing their movement time.

We also checked to ensure the variability measures and model fit parameters did not interact with the condition (results not reported), and instead tackle the two questions of error amplification and variability reduction separately.

5.5. Results

5.5.1. Error amplification

The ANCOVA of condition on movement time with pre-training movement time as a covariate revealed no significant effect of condition, *f*(1, 17) = .720; *p* = .409; η^2_p = .045 (Figure 5.5b).

The t-tests performed on the exponential curve fit parameters revealed no significant differences in the a parameter (an index of overall performance) between the Error Amp condition (M = 15.6, SD = 11.2) and the Control condition (M = 25.7, SD = 43.4); *t*(17) = .677, *p* = .51, η^2_p = 0.319; the b parameter (an index of rate of change of performance) between the Error Amp (M = -.462, SD = .224) and Control (M = -.620, SD = .476); *t*(17) = -.89, *p* = .39, η^2_p = -0.421; and finally the c parameter (an index of floor level performance) between the Error Amp (M = 1.49, SD = .162) and Control (M = 1.41, SD = .155); *t*(17) = -1.04, *p* = .31, η^2_p = -0.491.

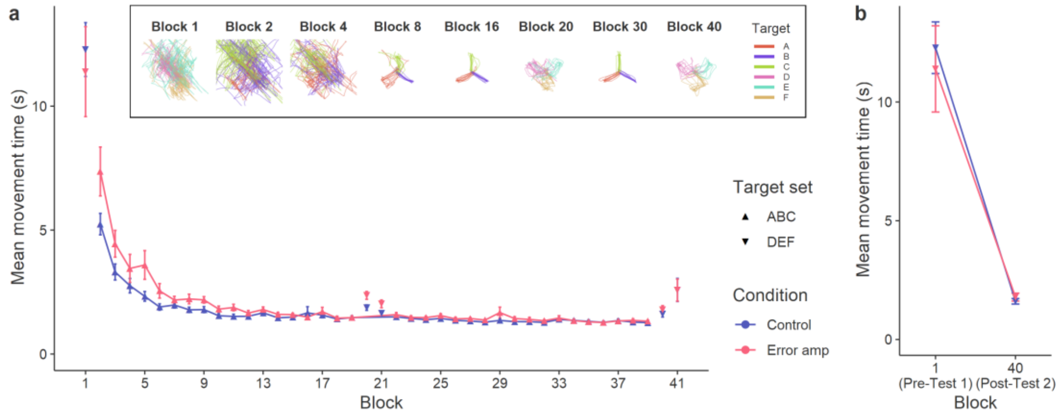


Figure 5.5 – Effect of error amplification on learning. (a) Both conditions show steep learning curves (reduction in movement time); participants in the Error Amp condition performed slightly worse in early training (though this was non-significant in terms of the coefficients of the exponential fit). (Note: Error amplification intervention was not applied for the assessment blocks, i.e. blocks with target set DEF.) Inset graph shows movements traces of a subset of blocks for the participant with median Block 1 movement time, highlighting typical movement patterns. (b) The change in performance across the two weeks was not significantly different between conditions.

5.5.2. Parameter space selection

The b parameter was significantly different from 0; *t*(16) = 4.31, *p* < .001, mean = -0.622, η^2_p = -0.622; indicating a reliable change in null-space variability measure over time (participant averages seen in Figure 5.6a).

The Spearman’s rank order analysis revealed no significant correlation between initial variability and performance; r_s(16) = .27, *p* = .279 (Figure 5.6c); but the final variability and performance were positively correlated; r_s(16) = .48, *p* = .044 (Figure 5.6d). Initial variability did not significantly predict training learning rate; r_s(16) = .22, *p* = .375 (Figure 5.6e). The correlation between variability rate of change and learning rate showed a positive relationship that was marginally significant; $r_s(15) = .5$, $p = .045$ (Figure 5.6f).

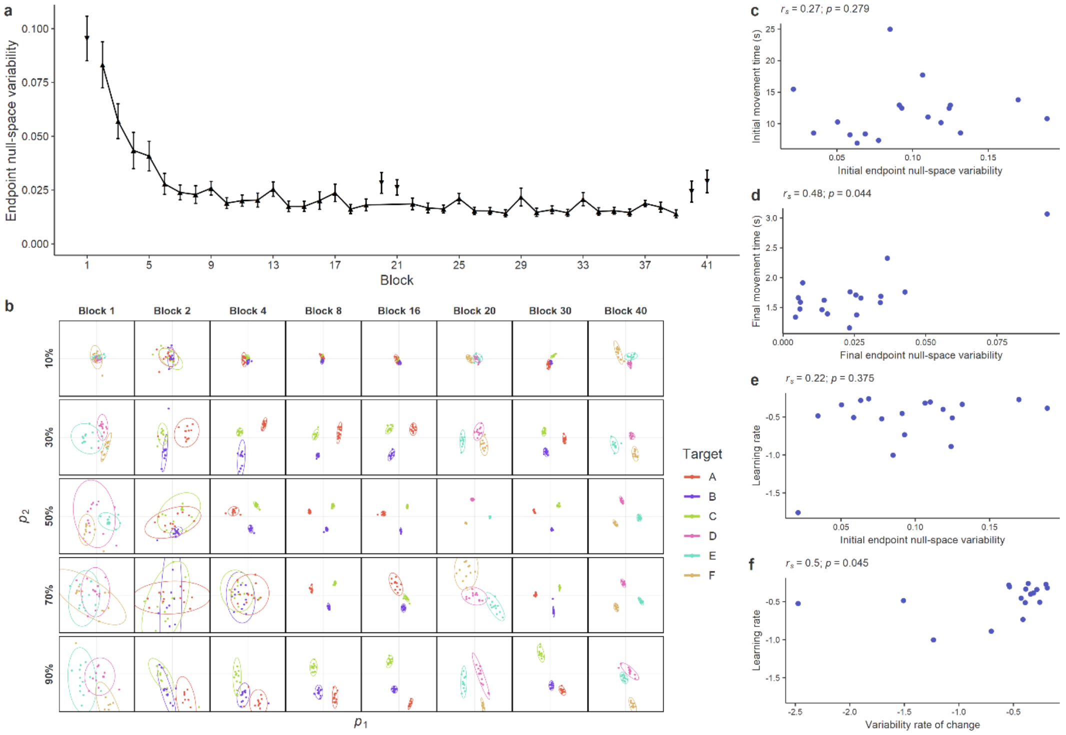


Figure 5.6 – Pose selection. (a) Endpoint null-space variability decreases over time. Points show the mean endpoint null-space variability across participants. Line connects training blocks. (b) The pose selection within the manifold (parameter subspace in which cursor meets the target) for 5 sample participants (those with 10th, 30th, 50th, 70th & 90th percentile Block 1 mean endpoint null-space variability) for a subset of blocks. Each target within each block have been fitted with a 95% confidence ellipsis (Fox and Weisberg, 2018). (c-f) Correlations of various variability and performance measures. Lower values of movement time indicate greater performance, and lower (more negative) values of the learning rate measure indicate a greater reduction in movement time (faster learning) [c] There was no relationship between the initial endpoint null-space variability and initial movement time, implying variability in itself is not a direct measure of performance. [d] There was a significant correlation between the final endpoint null-space variability and the final movement time, implying that after training, those with the ability to perform better also demonstrated consistency in their solution selections. [e] No significant correlation was found between the initial endpoint null-space variability and the learning rate, indicating that in this task, the variability at the outset is not related to change in performance. [f] A significant positive correlation was found between the rate of change of endpoint null-space variability and the rate of performance change indicating the participants who were able to reduce this variability the most were most able to improve their performance.

5.6. Discussion

We created a novel motor learning experiment to understand how variability manifests across the learning process. Recent research has brought into question the link between variability and learning (Braun et al., 2009; Dhawale et al., 2017; Cardis et al., 2017). Secondly, due to evidence of disruptive forces accelerating learning (Sigrist et al., 2013), this experiment investigated whether the artificial visual enhancement of error impacts on learning in the same way.

These data indicate that initial exploration of the solution space does not predict the subsequent learning rate of participants. Instead, the best learners were the ones who managed to simplify their degrees of freedom, by some process of elimination of redundant movements, thus reducing null-space variability. Contrary to our predictions, the error amplification intervention had no significant effect on learning, implying either an enhanced error signal cannot accelerate learning, or visual amplification of error is not a means of delivering an enhanced error signal.

A visual error amplification technique to manipulate error signal information was used in an attempt to understand the mechanisms of error-based learning. Previous studies have found evidence that the disruption of movements using force fields can enhance learning. The mechanisms of these effects are still to be understood, but one highly plausible explanation is that these force fields amplify the error signal and through correcting these errors the participants learn more quickly about how to successfully resolve the task. Here, we test this hypothesis by providing the stronger error signal without any haptic intervention, to examine its effects in an isolated experiment. We did not find any evidence that an increased error signal delivered though a visual error amplification intervention accelerates learning in our assessments. We also analysed performance measures during training, by fitting exponential curves to training data and performing t-tests on the resulting best-fit parameters. We found that across the 3 parameters (a, a measure of overall performance, b, a measure of rate of change of performance, and c, a measure of floor-level performance) there was no reliable difference between the amplification group and the control group, indicating that the error amplification intervention did not affect performance during training.

This experiment suggests that an increased error signal may not be a driving factor in studies reporting accelerated motor learning though interventions such as haptic disruption. Further experiments are required to examine the effects of error amplifying interventions to understand the mechanisms of motor learning.

We note that this experiment was not designed to enhance true error-based learning as the error signal presented to participant is non-veridical and thus, the motor correction for participants exposed to this group and the control should be equivalent. Further work is required to disentangle these processes in their contributions to motor learning. For example, one future experiment could look at facilitating error correction through other means aside from disturbance forces, which would help in understanding the mechanisms of the accelerated learning effect.

We aimed to capture the processes involved in solving the degrees of freedom problem in motor control, that is, finding solutions from redundant parameter spaces (Bernshteĭn, 1967). Many movements humans make every day are only one of any number of possible movements all equally adequate for performing the required task. Here, participants learned to solve a 2 degrees of freedom task (moving a cursor towards a target) using 4 degrees of rotations, hence a manifold of solutions existed for the participant to select for any given target. We were interested in how the variability in selection from this solution space (i.e. endpoint null-space variability) changed over time, and whether this had any bearing on an individual's learning rate.

First, we validated that this variability itself was not indicative of performance (Figure 5.6c). This is a key feature of this task, which assumes a range of solutions are valid, and a large variability in these solutions still allows for any level of performance. Then, we examined how the endpoint null-space variability was reduced over the course of training. We found stark reductions in this variability, with the analysis of curve fit parameters yielding a significant negative slope parameter (all participant saw a large decrease in movement time).

We also found a significant correlation between an individuals' learning rate and their variability reduction rate (Figure 5.6f), however since this is using data during training, we cannot be sure this effect is no driven purely by a model-free memorization process – where participants can quickly move their hands into a memorized pose after being presented with a given target, which would cause low movement times and low variability.

We found a significant positive relationship between final endpoint null-space variability and final movement time (Figure 5.6d). This does not imply any causal link but does show that those who managed to reduce their solution variability were also the participants who performed the best in this task. It is worth noting here that the data in question are those from the final assessment, which uses a different target set from those trained on. This was done so that any performance gain after training was a measure of the participant's understanding of the task dynamics, rather than a model-free memorization of a pose required to move to the target. The causal nature of this link should be investigated further – it is not known if the ability to perform consistent solutions (low endpoint null-space variability) is itself a cause of good performance, or if good performance is a cause of low variability. This task makes neither of these necessary, as it is possible to perform the task with high or low endpoint null-space variability without impacting performance. It has been previously observed that the redundancy in motor control is reduced through reduction of degrees of freedom (Newell, 1991; Li, 2006). The data here could be explained by this phenomenon, where those who managed to reduce their degrees of freedom, and thus the null-space variability, were able to perform better since the dimensionality of the search space has been reduced.

Finally, we examined weather an individual's initial endpoint null-space variability was predictive of the learning rate (Figure 5.6e). Exploratory analysis also revealed no significant correlations when using pre-post performance change or just final performance as the outcome variable; data not shown. This allowed us to investigate whether Wu et al.'s (2014) findings of initial variability in the task space predicting later performance changes would generalize to irrelevant variability (null-space). Singh et al. (2016) recently reported a significant relationship between a participant's initial null-space and subsequent learning rate in an adaptation task. We did not find any evidence to support these claims with no evidence of null-space variability in early training impacting on learning rates.

Informal observation of the solutions showed that there was no one-size-fits all best solution to each target (as designed), but instead a range of solutions were seemingly preferred across participants. This is seen by the different solutions used by the highlighted participants' block 40 data shown in Figure 5.6b. Subsequent work may focus on investigating the causes of these strategy selections, and how interventions might be able to encourage one type of solutio over another.

This work implies error amplification that may enhance performance (Hasson et al., 2016) has no benefit when presented on-line during the movement. It is speculated that perhaps the error signal is already very salient during online control and as such, the enhancement of this signal provides no additional benefit. Alternatively, it is possible that for this information to benefit the learner, there needs to be an explicit opportunity to utilise the enhanced error signal (Laura Marchal-Crespo et al., 2014) and this was not available to participants in this task. Finally, the solution variability element of this study expands upon an already established literature on null-space variability and the uncontrolled manifold concept (Scholz and Schöner, 1999; Scholz and Schöner, 2014;

Cardis et al., 2017). The methods of this chapter detail the mathematics of how to create a novel task with explicit redundancy, allowing the uncontrolled manifold concepts to be applied in a tractable manner and examined across the development of de novo skill acquisition.

5.7. Conclusion

We presented a novel motor learning task whereby participants attempt to solve a 2 degrees of freedom problem using 4 rotation degrees of freedom. The redundancy in the mapping of the 4 rotations to the 2D movement of the cursor creates a manifold of possible solutions available to the user for any given target position. This poses an abstract and novel challenge to the user, who has to learn to perform the task without any useful priors. We used this task firstly to manipulate the error signal presented to the participant during online control of the cursor when moving towards a target. In one condition, we amplified the error such that participants seemingly had a greater error to correct in an attempt to facilitate error-based learning. This was done to investigate whether the performance gains seen in disruptive error amplifying force interventions in other studies were the result of facilitation of a strengthened error signal. We found that visual error amplification had no effect on learning or performance when compared to a control condition with no intervention, providing strong evidence that increasing the error signal is not enough to facilitate the accelerated learning effect. The second part of this study examines how participants learn to perform skilled movements in a redundant system.

The task we created was designed such that there are a range of solutions all equally suitable for meeting task requirements, and we examined how the variability in these solutions changes over time. After verifying that this task irrelevant variability is independent of performance at the outset, we found that those who reduced this variability were the ones who performed the best. Future experiments should attempt to investigate a causal link between these, perhaps by manipulating this variability in a long-term learning setting.

6. Dissociating Selection and Execution Errors in Learning and Decision-Making

6.1. Overview

Sensorimotor error signals have played a key role in our examinations of motor learning to date in this thesis. In this chapter, we broaden our focus and across 3 experiments, we examine how these signals might serve to bias higher order cognition by focussing on decision-making.

Recent research indicates that, in two-alternative forced-choice sensorimotor decision-making tasks where participants are presented with options of equal value, but one yields a high rate of execution errors and another produces a high rate of selection errors, participants will systematically prefer to choose the option with a high rate of execution errors. In Experiment 1 we replicate these findings through a novel virtual reality two-armed action-bandit task. These biases could be accounted for by a recently proposed "movement-dependent" model of reinforcement learning, which predicts that execution errors attenuate value updating processes. However, a by-product of an incorrectly executed action is the uncertainty that arises from the lack of an opportunity to experience an outcome related to the selected choice.

Given that humans are information predators, a desire to reduce uncertainty could also account for these data. To disentangle these explanations, in Experiment 2, uncertainty was manipulated directly by asking participants to choose between two targets that yield the same rate of execution error, but one target also shows the counterfactual outcome (i.e. the outcome they would have received had the trial been executed properly). We reasoned that the introduction of a target that yielded counterfactual outcomes would reduce the uncertainty of the value estimate of that target and therefore must be selected less often (in comparison to a target that withheld this information on an execution error) if behaviour was driven by uncertainty reduction. No selection bias was observed between the two targets, indicating the uncertainty of a target’s value following an execution error does not facilitate a greater selection bias.

In Experiment 3, we examined how a greater association of selection and execution affected behaviour and found that a high association between selection and rate of execution errors drove participants to associate execution errors with the chosen target. Finally, we modified a movement-dependent reinforcement learning model, inferred fits of model parameters using Bayesian techniques, and compared models with Leave-Future-Out cross validation. There was no single best performing model for all experiments, but the parameter fits for the models broadly support the hypothesis of credit assignment gating under the assumptions of the models.

6.2. Introduction

When reward or punishment follows a series of actions, the brain has a credit assignment problem to solve. Specifically, it must infer the contribution of each individual action for the end result for future adaptive behaviour (Minsky, 1961;

Fu and Anderson, 2008; Wolpert et al., 2011). Consider for example the experience of losing a game of chess. Defeats in this context are typically the product of a series of actions that led up to the final move that ultimately terminated the game. Here, the chess player must assign credit to dozens of individual piece movements, negatively reinforcing those that were most responsible for facilitating a loss, such that they are less likely to be selected in subsequent chess games. Such problems of inference have been studied extensively in the computer science literature with artificial reinforcement learning agents (Minsky, 1961; Kaelbling et al., 1996), and across a range of topics in behavioural science- from association learning in rats (Mackintosh, 1975), to human motor learning (Berniker and Kording, 2008) and decision making (Fu and Anderson, 2008).

A recent twist on this classic credit assignment problem comes from a series of experiments investigating how value updating proceeds when the selected action plan may have been appropriate, but the agent fails to appropriate execute the planned action (McDougle, Boggess, et al., 2016; Parvin et al., 2018; McDougle et al., 2019). McDougle et al (2016) tailored a classic two alternative forced choice decision-making task, where participants selected between two “bandits” for rewards, with each bandit’s reward schedule varying in magnitude and likelihood. In an implementation of the classic formulation of the task, participants made selections between the bandits with keyboard button presses, with non-rewards clearly being signalled as the product of selecting the incorrect action plan (i.e. bandit). Here, participants showed the well-established phenomenon of risk aversion under uncertainty (Kahneman and Tversky, 1979)-

preferring to select bandits with high probability and low magnitude over riskier, low probability high magnitude options.

In a subsequent experiment, they introduced the probability that non-rewards could emerge from poor action execution by asking participants to make reaching movements towards the bandits. In these scenarios, end-point visual feedback was presented to participants to indicate whether the intended motor action was properly implemented. If the participant accurately reached the target, the bandit would change colour to indicate reward or non-reward – as in the classic version of the task. However, if the participant failed to hit the target, the participant would receive no reward and the feedback, showing a discrepancy between the reach and the intended target would indicate that the cause of this outcome was an “execution error”. In this version of the task, participants exhibited a striking reversal in selection strategy, showing a bias towards selecting low probability, high reward bandits. The authors proposed that this phenomenon may be accounted for participants discounting (or “gating”) when updating the value estimation of the selection, when the “blame” for an error could attributed to the sensorimotor system (McDougle, Boggess, et al., 2016). In other words, while selection errors provide information about the intrinsic value of the target, no value can be inferred following an execution error, given the non-reward can be attributed to a poorly executed motor plan.

Follow up studies directly contrasting targets with differing degrees of execution error and selection error, but with equivalent expected value, have shown that participants are systematically biased towards selecting options that have a higher likelihood of eliciting execution errors (McDougle et al., 2019; Mushtaq et al., 2019). A potential explanation for this bias comes from Parvin et al., (2018) who, in ruling out the hypothesis that gating was driven by bottom-up sensory prediction signals, found that manipulating the participants' top-down belief in their ability to influence the outcomes ("agency") modulated the extent to which participants gated value following execution error. In generalising this line of reasoning, this process of gating may be driven by the fact that execution errors have properties that provide information about the correctability of subsequent behaviour and thus, participants are drawn towards making these corrections c.f. a selection error. This has been modelled as a "persistence" parameter (McDougle et al., 2019), where we assume some extra value arising from re-selecting the same target following a miss.

This sense of agency, however, is not the only difference between a selection and execution error and other intrinsic information properties of these difference outcomes may contribute towards this bias. For example, whilst it is clear that selection and execution error outcomes yield no reward, in contrast to selection errors- where one learns that the chosen course of action would not, and did not, produce a reward- execution errors leave open the possibility that the selected option may have yielded a reward, *if only* the action was properly implemented.

Over an extended period of time, a selected action that elicits a high proportion of execution errors will lead to a less certain estimate of the option's value relative to an outcome that provides feedback. So why might participants be so inclined to continue selecting this option? An explanation may be rooted in an exploration-exploitation trade-off, with humans exhibiting short-term desires to minimise uncertainty (explore) as well as maximise reward (exploit), as an optimal means of maximising long-term gain (Cohen et al., 2007). In this way, an option with a high chance of obtaining a reward may intentionally be avoided while an agent seeks out an alternative with a much less certain reward probability.

This type of strategy seems to facilitate several distinct patterns of behaviour, depending on the context. Whilst there is no one-size-fits-all solution for optimal exploration and exploitation (compounded by the fact that real-world environments are non-stationary), humans are sensitive to changes in reward. When rewards are scarce following one course of action, there is a tendency to switch to an alternative (Daw et al., 2006). However, depending on the task and its context, the contrary is also common, where humans try harder at the same action when reward is reduced, rather than exploring other options (Rabbitt, 1966; Laming, 1979; Gratton et al., 1992; Cohen et al., 2007). Taking into account the tendency of organisms to reduce uncertainty, we posit that an alternative account of the results reported to date, may be driven by a desire to reduce uncertainty, which acts as an attractor towards high execution error targets.

To test these ideas, a novel Virtual Reality based 2-armed bandit task was created, where participants selected between targets for reward. Trials that failed to elicit a reward where either presented as errors arising from incorrect selections or poorly executed actions. While both bandits resulted in the same amount of reward, one bandit systematically elicited more execution errors relative to selection error whilst the other elicited more selection errors relative to execution errors. In Experiment 1, previously reported results showing participants gravitate towards targets with higher frequency of execution errors were attempted to be replicated. In Experiment 2, the contribution of uncertainty in driving this behaviour was examined. Here, counterfactual outcomes were introduced with one target also yielding “fictive” outcomes where participants were shown whether the selected action would have produced a reward if it was executed correctly. If participants’ behaviour in these tasks is the product of a need to resolve uncertainty, then participants would be biased towards the target that yields fewer “fictive” outcomes.

We also examined how providing certainty about the value of the selected action even in the presence of an execution error affects the apparent gating in the credit assignment process. The gating effect (attenuated value updating during credit assignment) may arise from either the presence of an execution error, or the uncertainty that arises on an execution error when the reward information is hidden. Previous experiments have not de-coupled these two features. This could be investigated through examining behaviour following a fictive outcome compared to regular miss outcome, such as a change in value estimate or differing selection behaviour.

Finally, we asked whether participants would continue to discount selection errors even if they were closely tied to the selection. . When execution errors are clearly tied to their respective selected actions, we predict action execution errors will be treated almost as if it were a selection error. The experiment tests various levels of the extent to which execution errors are tied to action selections, by presenting the participant with two targets of either very similar or very different execution error rates. In other words, we ask if one keeps selecting an action that systematically produces an execution error, at what point does this become a bad choice? The behaviours that execution errors and selection errors elicit are expected to converge in the case where the two are highly associative.

6.3. Methods

6.3.1. Action Bandit Task

We adapted a classic “bandit” task, often used to study how humans learn the reward probabilities of several independent systems (Daw et al., 2006). Using a virtual reality head-mounted display system, participants were asked to select one of two spherical bandits by making “swiping” actions (without online feedback) towards them using a controller. On successfully swiping the target, the bandit would either, (a) open to reveal a star, earning the participant 1 point (a reward trial), or (b) open and reveal that the bandit was empty, and the participant received 0 points, a non-reward indicating that the participant made a reward prediction error (RPE). If participants failed to accurately swipe the bandit, they would receive no reward and end-point feedback indicated that this non-reward was the product of an execution prediction error (EPE). Participants were instructed to choose the target they believed had “the highest chance of giving them a star, at that moment, based on their prior experience with the targets".

Each trial began with a movement of a white spherical cursor (diameter 1cm; controlled via a hand controller held in the participant's preferred hand) to a start point (Figure 6.1). After a short delay (sampled from a uniform random distribution of 500-800 ms), the starting point turned green, a "whistle" sounded, and the two targets appeared. The participant then had up to 1500ms to move from the start point, and a further 600ms to attempt to swipe a target (starting from when they exited the start point), allowing time for a decision to be made but requiring a single, fast action. Participants who moved too slowly were shown an error message with a buzzer sound and attempted the trial again. The two targets (diameter 4 cm) were positioned on a "ring" of radius 25cm at 190° and 350° from the horizontal and were randomly assigned colours magenta or yellow. Both the target positions and colours were randomly assigned. The participant had no vision of their hand position throughout the movement and could only see the position of their cursor after they completed the movement (shown stationary on the edge of the ring). The measurements here are for a participant with their height equal to the reference height (170 cm). Participants' heights were measured using the reported HMD position, and task scale was multiplied by the ratio of the participant height to the reference height, making the ring, targets, and cursor relatively bigger or smaller. The range of heights was 129cm to 178cm, which meant the task was scaled between 76% and 105% of this reference size depending on the participant height. This scaling, determined through pilot testing, was performed to accommodate for a variety of participant heights and arm lengths and ensure each participant could comfortably complete the task.

Successfully hitting the target was made clear by including a “sword slicing” sound, a “sparks” particle effect, as well as a 200ms vibration in the held controller at the instance the participant’s cursor hit through the ring. Missing the target had no associated sound, particle or haptic feedback effects. On hitting the target, the target would split in half, sometimes revealing a reward (a golden star) worth 1 point. The star then moved towards the participant’s body to that it had been collected, and 1 point is added to the participant’s score on the instruction board, playing a pleasant “ding” sound. The instruction board housing the current score is shown throughout, 1.3m away from the participant. Additionally, text feedback was shown on the scoreboard, with instructions “Collect the stars” which changes to appropriate messages depending or the outcome, or to error messages when the participant didn’t follow instructions (e.g. moved too early, moved too slowly). The task software was developed with Unity 2018.1 (Unity Technologies, 2018), the SteamVR SDK (Valve Corporation, 2018), and the Unity Experiment Framework (Brookes et al., 2019). The Psychology Research Ethics Committee at the University of Leeds approved the research.

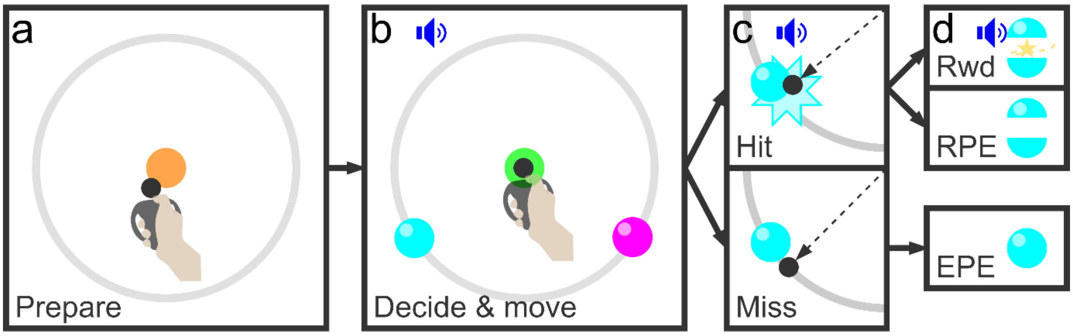


Figure 6.1 – Example trial and three possible outcomes. (a) Participants move their hand towards a red start marker. (b) When it turns green, a whistle sound is played, and the two targets appear. Participants must choose either target by swiping their hand through it. (c) The chance of hitting or missing the target is predetermined, and with audio, visual and haptic effects playing on a hit, with no feedback on a miss. (d) If the target was hit, the target splits into two hemispheres and reveals either a star which travels towards the player and emits a “ding” sound (Reward outcome: 1 point) or nothing (RPE outcome: 0 points). On a miss, the target remains closed (EPE outcome).

6.3.2. Outcomes

The three outcomes used here (Reward, RPE, EPE) had their probabilities fixed (and therefore all outcomes pre-determined), and pseudo-veridical feedback was implemented to ensure outcomes appeared genuine where possible. The pseudo-veridical feedback was implemented by offsetting the position of the cursor on the ring only when necessary, that is, when the predetermined motor outcome (hit or miss) differed from the participants actual motor outcome. If the participant missed the target on a predetermined hit trial, the cursor position was offset to show it touching the target plus a randomly generated offset angle (0.1-0.3°) towards the target. If the participant hit the target on a predetermined miss trial, the cursor position was positioned to be touching the target plus a randomly sampled offset angle (0.1-3.0°) away from the target. In both cases, the same direction of the error was maintained. Where the predetermined and actual motor outcomes matched, the cursor position was shown at the actual position it hit the ring. Movements that were too far from either target (error of more than

30°) or were out of bounds (moved the hand more than 15cm forwards or backwards from the ring) were met with the error buzzer sound, and the trial was repeated.

6.4. Experiment 1: Manipulating Error Type

In Experiment 1, the probabilities of the two targets were designed so that they had an equivalent expected value with equal amount of reward but differed in the frequency of the non-reward outcome feedback type. The EPE+ target yielded more execution errors, whilst the RPE+ target yielded more selection errors (Table 4). Each participant completed a total of 350 trials over approximately 25 minutes. An example single trial is shown in Figure 6.1.

Table 4 – Outcome probabilities for the two targets in Experiment 1

	EPE+ target	*RPE+ target*
Reward	30%	30%
RPE	20%	50%
EPE	50%	20%

These target probabilities allow us to determine how participants treat the two different types of error (RPE and EPE) facilitate both long-term behaviour of the overall selection preference, and short-term behaviour of reselecting the same target (or switching to the alternative) following these errors. The goal of this study was to examine whether previously reported results (McDougle, Boggess, et al., 2016; McDougle et al., 2019), which show participants being biased towards selection errors over execution errors, could be replicated in this new virtual reality task.

The second goal was to examine how likely participants were to reselect the same target following each types of outcome on a trial-by-trial basis. The “gating” movement-dependent RL model (McDougle, Boggess, et al., 2016) does not explicitly predict these short term behaviours. Instead, this model implicitly suggests a preference for the EPE outcome over the RPE outcome, and that Rewards should be preferred over EPE outcomes. One might expect these preferences to be apparent in the reselection rates of a target following each outcome, with a higher reselection rate indicating preferences of the outcomes.

6.4.1. Participants

Twenty-two participants (age range 18 - 29 years, mean = 19.3 years; 18 right hand dominant; 4 female 18 male) were recruited from a 1st year undergraduate computing class at the University of Leeds and were each paid £5 for their participation.

6.4.2. Statistical analyses

To measure target selection bias, we calculated the percentage point (p.p.) difference between the overall selections of the two targets. A one-sample t-test was performed comparing the participant’s overall selection biases to 0, and the significance level was set at $\alpha = 0.05$.

Controlled reselection rate presents a measure of how likely a participant was to reselect the same target following each outcome relative to an individuals' own reselection rate. This was calculated by first computing the percentage of trials where the participant reselected the same target following each outcome for each target and subtracting from this the percentage of trials where the participant reselected the same target (regardless of outcome). Here, a within-subjects ANOVA was performed, with reselection rate as the outcome variable, and the previous trial outcome, previous trial target, and the interaction between these two were used as independent predictors. All data met assumptions of normality through assessment by histogram, Q-Q plots, and Shapiro-Wilk tests. Trials where a participant failed (e.g. moved too slowly, swiped too far from either target) were not included in any analyses.

6.4.3. Results

We found a statistically significant selection bias towards the EPE+ target in line with previous findings; *mean* = +17.8 p.p. bias towards EPE+, $t(21) = 5.18$, $p < .001$, $d = 2.209$ (Figure 6.2a).

The reselection rate analysis revealed a significant main effect of the previous trial outcome on the reselection rate; $f(2, 21) = 39.28$, $p < .001$, $\eta^2_p = 0.317$. The main effect of target approached statistical significance; $f(1, 21) = 3.55$, $p = .062$, $\eta^2_p = 0.020$; there was no significant interaction; $f(2, 21) = 2.10$, $p = .13$, $\eta^2_p = 0.023$ (Figure 6.2b). Pairwise comparisons of the mean reselection rate across three outcomes (collapsed across targets, Holm corrected) revealed reselection rate to be significantly higher following Reward compared to RPE ($p = .004$), and significantly lower compared to EPE (*p* = .009). Reselection rate following the EPE outcome was also significantly higher than following RPE (*p* < .001).

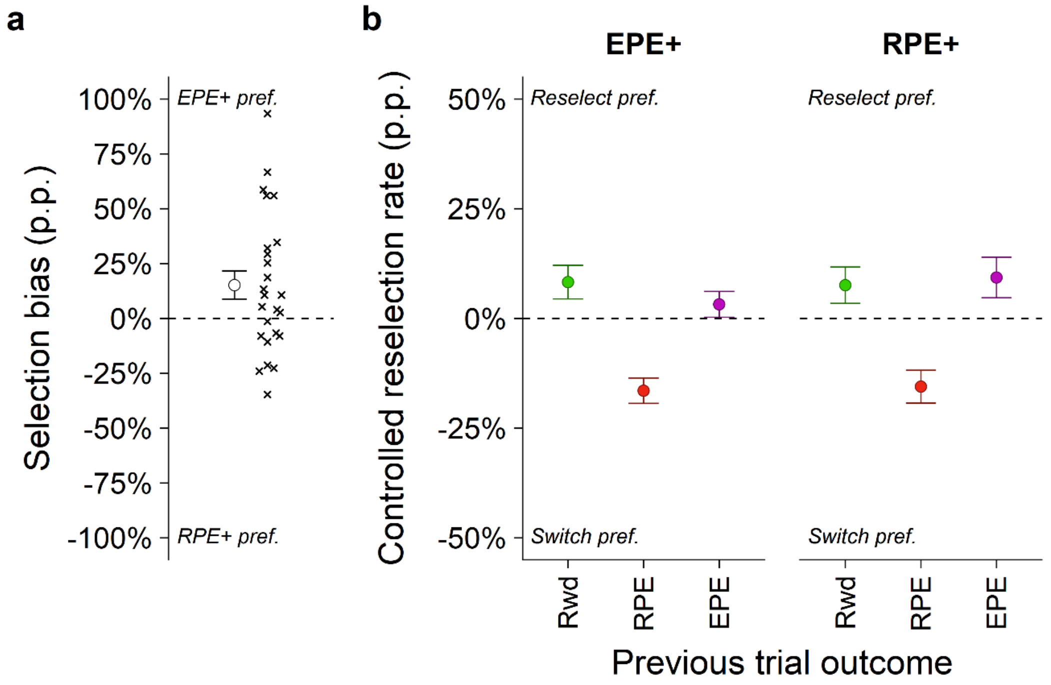


Figure 6.2 – (a) There was a statistically significant selection bias with a preference towards the EPE+ target in Experiment 1. (b) There were statistically reliable differences in reselection rate across trial outcomes. Error bars show S.E.M.

The results here are largely consistent with the hypothesis that the negative value associated with an EPE outcome attenuated in the target value updating process. Interestingly, Figure 6.2b reveals a significantly higher rate of reselection of the same target following an EPE. This may be due to a short-term “persistence” effect that facilitates reselection following an uncertain reward or the execution error signal (discussed in Discussion).

6.5. Experiment 2: Reducing Uncertainty Through Fictive Outcomes

In Experiment 2, the task was modified to allow participants to experience “fictive” outcomes when execution errors were made i.e. indicate what would have happened if the participant had executed their action accurately. Specifically, in the case of an execution error, the target would still open to reveal an empty balloon or star inside. Critically, revealing of the star here did not award the participant 1 point and this was made clear by having the star fade away with an “electrical power down” sound, coupled with no increase to the participant’s cumulative score. Additionally, the salient feedback presented on a hit (see section: Action Bandit Task) presented on a hit, vs the lack of this feedback on a miss, made the differences between a real and fictive outcome very apparent.

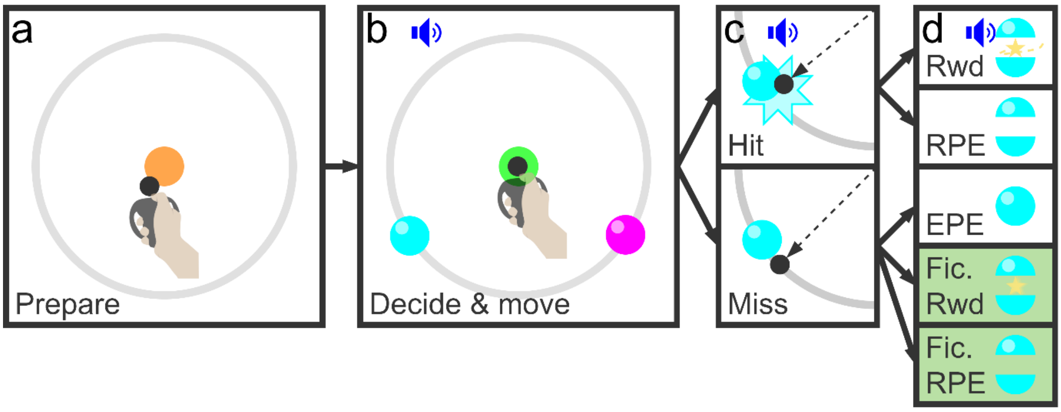


Figure 6.3 – The trial procedure is identical to Experiment 1. (a) Participants move to the centre point. (b) Participant choose one of the two targets by swiping through with their hand. (c) Participants are told they either hit or miss. (d) On hit, the target opens to reveal a star (1 point), or nothing (0 points). On missing, the target could stay closed, and no points are earned. In Experiment 2 specifically, one of the targets has the possibility of yielding a fictive outcome (shown here shaded green), where the target opens to reveal, following a miss, either a star which proceeds to fade away (Fictive Reward outcome – still earning 0 points), or nothing (Fictive RPE outcome).

The two new outcomes (Fictive Reward, Fictive RPE), brings the total to five possible outcomes (Table 5). Experiment 2 was reduced to 300 trials to allow for time to explain these additional outcomes within the same time constraints as the previous Experiment.

Table 5 – Experiment 2 introduces two fictive outcomes which show selection information upon (seemingly) unsuccessful actions.

		Selection		
		Successful	*Unsuccessful*	*Unknown*
Action	*Successful*	Reward	RPE	-
	Unsuccessful	Fictive Reward	Fictive RPE	EPE

🟧	Experiment 1, 2 & 3
🟩	Experiment 2 only

In Experiment 1, the two targets differed only in the frequency of the two types of errors (RPE vs EPE). As a side effect, an execution error outcome also could facilitate a lower level of certainty about the target. The two targets presented in Experiment 2 were selected to only manipulate the level of certainty of each of the two targets (both targets had an equal rate of hitting and missing, and an equal overall expected value). One target, "Certainty-" acted in exactly same way as the targets did in Experiment 1, i.e. on an execution error, the participant was given no knowledge of whether their selection would have yielded a reward if they did hit the target (the EPE outcome). The "Certainty+" target differed in that on an execution error, it always revealed the fictive reward (or lack thereof) that would have been awarded if the participant had correctly executed the action (Fictive Reward and Fictive RPE outcomes). It follows then that over time, the participant's value estimate of the Certainty+ target would therefore be more precise than that of the Certainty- target.

Table 6 – Outcome probabilities for the two targets in Experiment 2

	Certainty+	*Certainty-*
Reward	30%	30%
RPE	20%	20%
Fictive Reward	30%	-
Fictive RPE	20%	-
EPE	-	50%
Open (total)	100%	50%
Closed (total)	-	50%

This experiment tests the hypothesis that the apparent risk-seeking behaviour of a selection bias towards choices that yield more EPE outcomes is driven (in part) by a desire to reduce uncertainty about the targets. The target that yields more EPE outcomes often "hides" its true value from the participant, and as such, would require a greater number of visits to be certain about that value. The new outcomes also allowed us to disentangle the different aspects of any preference shown towards the EPE outcome alone, regardless of the target. Experiment 1 compounds the lack of information and the motor error into a single outcome, EPE. In Experiment 2, we can see the contribution of these two features on the reselection rate separately.

6.5.1. Participants

Thirty-one participants were recruited to the study and were each paid £5 for their participation. 3 participants were excluded from analyses as their overall absolute target selection bias was greater than 90%, indicating that they did not follow task instructions. The final number of participants was 28 (age range 18 – 25 years, mean = 19.2 years; 28 right hand dominant; 21 female, 7 male).

6.5.2. Statistical analyses

The same statistical tests performed for Experiment 1 were also performed here (t-test on selection biases, ANOVA on reselection rates). However, the reselection rate analysis was performed only with the Certainty- target (since this target yields the same three outcomes as Experiment 1). In addition, a two-way ANOVA of reselection rate of the Certainty+ target, with previous outcome type (Fictive vs Regular) and previous outcome reward (Reward vs RPE) as independent predictors, was performed. This allowed us to discriminate between the impact of fictive outcomes and reward independently. All data met assumptions of normality through assessment by histogram, Q-Q plots, and Shapiro-Wilk tests.

6.5.3. Results

Experiment 2 tests the hypothesis that the tendency to select targets that yield EPEs over RPEs is in-part caused by a desire to reduce uncertainty about both targets. This hypothesis predicts a selection bias towards the Certainty- target over the Certainty+ target, which remains closed following a miss, and opens producing fictive outcomes following a miss, respectively. We found no statistically reliable bias, with the mean selection in fact being biased towards the Certainty+ target; *mean* = +5.0 p.p. bias towards Certainty+, *t*(27) = 1.72, *p* = .10, *d* = 0.651 (Figure 6.4b).

The reselection rate analysis revealed that, regarding the Certainty+ target, there was a main effect of the presence of a reward (regardless of if it was fictive); *f*(1, 27) = 10.21, *p* = .002, η^2_p = 0.051; outcome type (fictive vs regular) showed no significant effect; *f*(1, 27) = 0.31, *p* = .57, η^2_p = 0.002; with no significant interaction; *f*(1, 27) = 0.02, *p* = .89, η^2_p < 0.001. For the Certainty- target, there was a main effect of outcome on reselection rate; *f*(2, 27) = 3.40, *p* = .041, η^2_p = 0.037. Pairwise comparisons of the mean reselection rate across three outcomes (collapsed across targets, Holm corrected) revealed no significant difference between any pair; Rwd vs RPE, *p* = .061; Rwd vs EPE, *p* = .890; RPE vs RPE, *p* = .061.

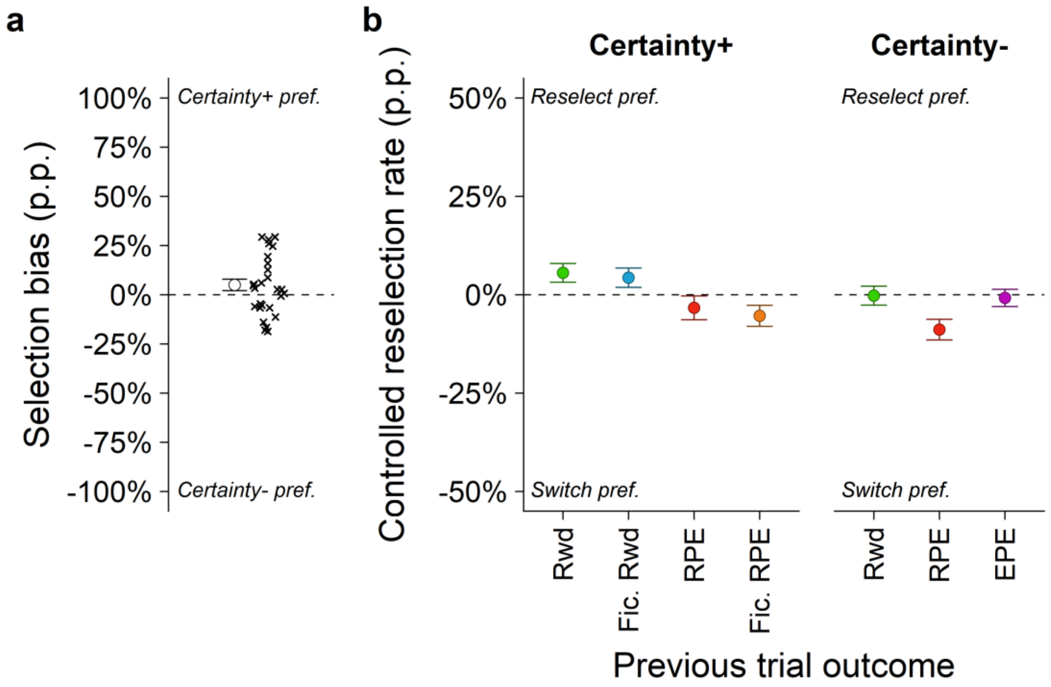


Figure 6.4 - (a) There was no statistically significant selection bias towards either target in Experiment 2. (b) There were statistically reliable differences in reselection rate between reward. Error bars show S.E.M.

The reselection rates on the Certainty- target following each outcome imply there is no reliable difference in preference of the outcomes relative to their fictive counterparts (Rwd vs Fic. Rwd, RPE vs Fic. RPE). Interestingly, the EPE outcome in the Certainty- target does not sit in between the Reward and RPE reselection rates, but is closer to the Reward reselection rate.

6.6. Experiment 3: Manipulating Execution Error – Selection Coupling

In addition to differences in uncertainty, another fundamental difference between an RPE outcome and an EPE outcome is that RPE appears to signal information about the property of the target while an EPE is a property of one's own action. In the VR-bandit task, there is no visual information at the start of a trial that might indicate one target is harder to hit than another (e.g. both targets have an identical size) and as such, there is little reason for them to start the experiment by believing the EPE to be a feature of the target. We hypothesised that if an EPE outcome was to be presented as an apparent property of the target (i.e. the selection we made impacted the miss chance) it would be treated the same as an RPE and indexed by the same reselection rates.

A key feature of an EPE (and a potential driver for the bias towards it) is that it is correctable by the participant. If the lack of correctability becomes apparent, we expect EPE outcomes to elicit low levels of target reselection. We designed Experiment 3 to test this hypothesis. Here, three different sets of targets with different levels of coupling between execution error and selection (Low – Medium – High; Table 7) were used in a between-subjects design experiment. The diverging differences in probability are a means of presenting an implicit coupling between target selection and rate of execution errors. A high coupling of target selection and execution errors is hypothesised to then prohibit the gating mechanisms, since gating presumable happens when the observed feedback is unrelated to target value. Here, execution errors were presented as coupled with target selection. To achieve this, only difference between the two targets in each case was the probability of missing (EPE). The probabilities of Reward and RPE are set such that a hit always yields 50% chance of a point.

Table 7 – Target probabilities for the three groups in Experiment 3 (rounded to 1.d.p). The three groups had different levels of execution error – selection coupling, implemented by a 15, 30, or 60 percentage point difference between the targets.

	Low coupling		***Medium coupling***		***High coupling***	
	EPE+	*EPE-*	*EPE+*	*EPE-*	*EPE+*	*EPE-*
Reward	21.3%	28.8%	17.5%	32.5%	10.0%	40.0%
RPE	21.3%	28.8%	17.5%	32.5%	10.0%	40.0%
EPE	57.5%	42.5%	65.0%	35.0%	80.0%	20.0%

We also used this opportunity to examine the impact of learning on decision-making. In this trial configuration, in contrast to the previous experiments, expected value differences between targets were introduced. Here, EPE+ targets led to a lower overall value (which we expect to drive behaviour towards the EPE- target in all cases). The gating movement-dependent RL model implies that as participants experience the two targets in Experiment 1, the value gating facilitated by the EPE outcome leads to a belief that the Experiment 1 EPE+ target has a higher value than the RPE+ target. We use Experiment 3 to compare the behaviour elicited from real value differences to a belief or perception of differences in value predicted by Experiment 1. A trial in this study was the same as Experiment 1 (Figure 6.1). Participants performed 150 trials to meet a shorter time constraint, data from Experiment 1 revealed selection biases and reselection rate differences arising after a smaller number of trials than the 300+ used previously (data not shown).

6.6.1. Participants

Participants in this study were awarded 1 credit in the University of Leeds Undergraduate Participant Pool Scheme, and additionally were entered into a pool to win one of 6 prizes (£30 x 4, £20, £10) – with the winners chosen based on “a combination of number of points and movement accuracy”. Participants that did not experience all three outcomes for each target (due to very large selection biases from the outset) were excluded from analyses. This left 26, 28 and 26 participants in the low, medium & high coupling groups respectively. The low coupling group had an age range of 18 – 24 years, mean = 18.9 years; 26 right hand dominant; 25 female, 1 male. The medium coupling group had an age range of 18 – 25 years, mean = 19.5 years; 27 right hand dominant; 28 female. Finally, the high coupling group had an age range of 18 – 22 years, mean = 19.1 years; 24 right hand dominant; 25 female, 1 male.

6.6.2. Statistical analyses

For Experiment 3, a t test was performed on the selection bias for each condition. In addition, three separate ANOVAs (one for each trial outcome) were performed with reselection rate as the dependent variable, and previous trial target and coupling level (low, medium or high) as independent variables. This would reveal if the increasing execution error – selection coupling affects the relationship between reselection and target, for each of the outcomes (Reward, RPE, EPE). All data met assumptions of normality through assessment by histogram, Q-Q plots, and Shapiro-Wilk tests.

6.6.3. Results

Experiment 3 consisted of three groups with diverging levels of probability of EPE outcomes between targets. This resulted in a scenario where there was a real value difference between the targets. The t-tests for each condition revealed a no significant selection bias in low coupling condition; *mean* = 2.7 p.p. bias towards EPE-, *t*(25) = -0.51, *p* = .61, *d* = -0.202; but a significant bias in medium coupling; *mean* = 15.0 p.p. bias towards EPE-, *t*(25) = -2.40, *p* = .024, *d* = -0.940; and high coupling; *mean* = 33.5 p.p. bias towards EPE-, *t*(25) = -4.77, *p* < .001, *d* = -1.872 (Figure 6.5).

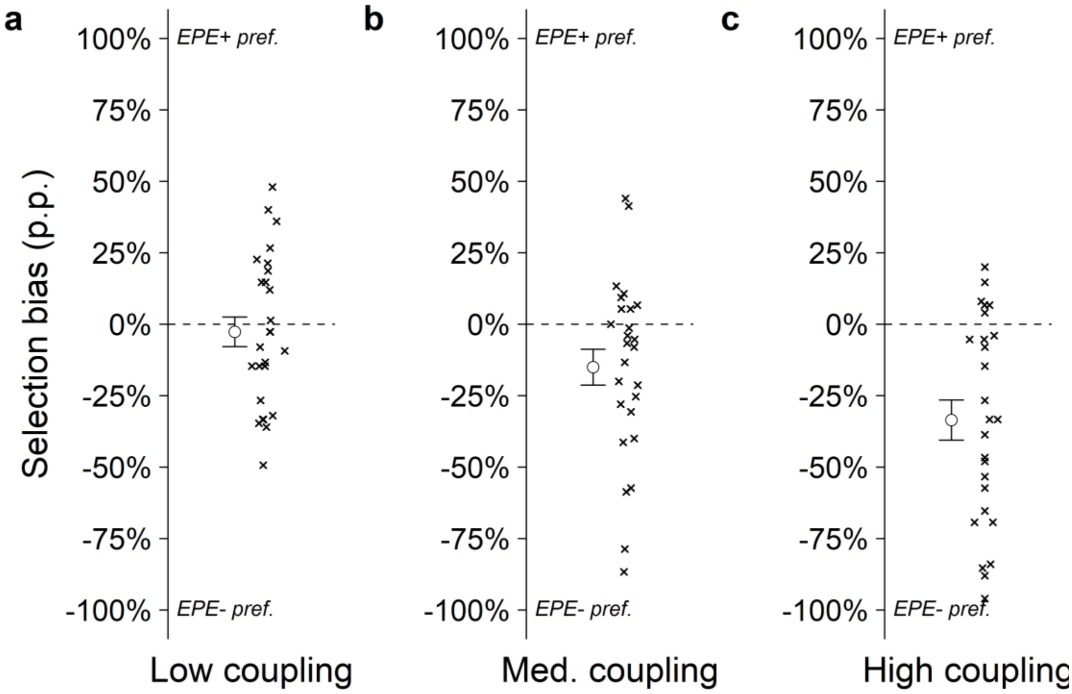


Figure 6.5 - Selection biases for Experiment 3 (a) low coupling, (b) medium coupling and (c) high coupling. Results show a significant selection bias at medium and high coupling level with a larger effect size at high coupling.

For the Reward outcome, the ANOVA revealed no statistically significant main effect of previous target on the reselection rate; *f*(1, 70) = 2.398, *p* = .13, η^2_p = 0.006; or coupling level; *f*(2, 70) = 0.655, *p* = .52, η^2_p = 0.015; and no significant interaction; *f*(2, 70) = 0.806, *p* = .45, η^2_p = 0.004. For the RPE outcome, there was a main effect of previous target; *f*(1, 70) = 5.772, *p* = .019, η^2_p = 0.027, with the EPE- target facilitating an overall greater chance of reselection; and marginal effect of coupling level; *f*(2, 70) = 3.128, p = .05, η^2_p = 0.055; but no significant interaction; *f*(2, 70) = 1.355, p = .45, η^2_p = 0.013.

The EPE ANOVA also indicated a main effect of previous target; *f*(1, 70) = 30.641, p < .001, η^2_p = 0.147, again with the EPE- target facilitating a greater level of reselection; but not of coupling level; *f*(2, 70) = 0.208, p = .81, η^2_p = 0.003; the interaction here was significant; *f*(2, 70) = 5.469, p = .006, η^2_p = 0.006. (Figure 6.6). The interaction in the EPE case can be seen as a steep reduction in reselection rate at higher coupling levels, exclusively in the EPE+ target.

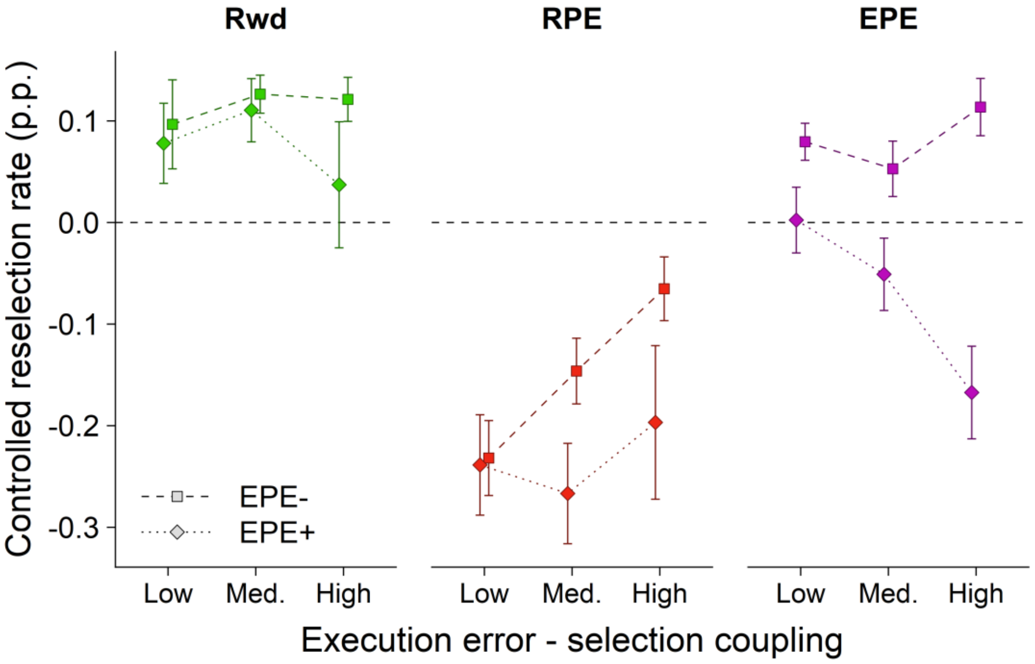


Figure 6.6 – Relationship of reselection rate vs outcome and target varies with coupling level. Reselection following a reward was not statistically significant across targets and coupling level, but there were statistically significant effects of target and target-coupling level interaction on reselection following either RPE or EPE outcomes.

6.7. Modelling

We tested five classical error-based learning models (Rescorla and Wagner, 1972) adapted from the gating model proposed in McDougle et al., (2016). All assume a model-free mechanism of learning, whereby the value of each target is updated using a TD error, and probability of selecting each target is related to the relative value estimates of each target. In these models, the TD error δ on trial t is the difference between the current value estimate of chosen target x ($V_t(x)$) and observed reward on trial (r_t),

$$\delta_t = r_t - V_t(x). \qquad 15$$

The observed reward r has a value of 1 or 0 when the target reveals a star or nothing respectively (even on a miss, i.e. the Fictive Reward and Fictive RPE outcomes in Experiment 2), and 0 on an EPE outcome. Then, the value estimate for the chosen target x is updated using the TD error and an outcome-dependent learning rate η^*,

$$V_{t+1}(x) = V_t(x) + \eta^* \delta_t. \qquad 16$$

Crucially, η^* is selected based on whether the outcome was a hit ($h = 1$) or miss ($h = 0$), simulating "gating" of credit assignment using gating coefficient k on miss trials,

$$\eta^* = \begin{cases} \eta & \text{if } h = 1 \\ k\eta & \text{if } h = 0 \end{cases}. \qquad 17$$

Where η the base learning rate. On each trial, the model assumes probabilities of selection are then derived from the value estimates of the two targets using a SoftMax function with inverse temperature parameter β. For example, the probability P of selecting Target 1 on trial t is

$$P_t(1) = \frac{e^{\beta V_t(1)}}{e^{\beta V_t(1)} + e^{\beta V_t(2)}}. \qquad 18$$

This formulation will be referred to as the *basic gating* model from hereon in as subsequent competing models involve small modifications to this foundation model.

The *miss gating* model is relevant only Experiment 2 and assumes that the reward r is always 0 on a miss, even when the target opens to reveal a missed star. In effect, this assumes participant ignores the fictive outcome, treating a miss as 0 reward.

The *split gating* model assumes the gating rate used in the learning rate selection (Equation 17) can be different for either target ($k(x)$), testing the hypothesis that the gating rate is based on participant’s belief that the EPE outcome is selection dependent.

The *shift gating* model assumes the gating rate k is not fixed in time, but is trial dependent, and linearly shifts from k_{Start} to k_{End} over time (between the first and last trial of the session, respectively).

Finally, a "*miss reward*" model was tested which did not have any gating mechanism (i.e. $k = 1$), but instead assumes a subjective reward r^* instead of just the target reward r is used to generate delta value estimates (Equation 15). Specifically, when participants miss, $r^* = r + r_{miss}$, or $r^* = r$ otherwise. This models the idea that a miss provides some additional intrinsic reward, e.g. valuable motor information that can be used to improve future executions.

While maximum likelihood estimation methods could have been used to estimate model parameters, we opted to use Bayesian estimation techniques to infer distributions of possible parameter values in each model. We estimated the posterior distribution $P(\text{Parameters}|\text{Data})$ using the No-U-Turn algorithm (Hoffman and Gelman, 2011) implemented RStan 2.18.0. The gelman-rubin statistic ($\hat{R}$) (Gelman et al., 2014) which assesses convergence was well below 1.1 for all parameters. Hierarchical implementations of these models with parameter estimates per participant yielded inconsistent fits which failed to converge (high $\hat{R}$ values) in a reasonable number of samples, so we opted to estimate a single value of each parameters for each experiment and use the 95% highest density intervals (HDI) to assess significance. The HDI of each parameter in the posterior provides an upper and lower bound which has a 95% probability of containing the true parameter value. HDIs were estimated using the methods outlined in Hyndman (2002) implemented as the hdrcde package for R. Eight chains of 4,000 samples were taken from the posterior; the first half were discarded as warmup samples.

The values of η were constrained between 0.1 and 1.0 (avoiding a local minimum encountered at $\eta = 0$ in some chains), β values were constrained between 0.001 and 100, and the model specific constants (k for gating models, r_{miss} for miss reward model) were constrained between 0 and 1. The initial value estimates were assumed to be 0.5 for both targets. Parameter priors were uniform across these constraints.

We compared each incarnation of the gating model to assess the predictive power of each model. We opted to use the estimated Leave-Future-Out Cross-Validation (LFO-CV; Bürkner, Gabry, & Vehtari, 2019) technique to estimate the ability of each model to predict out-of-sample observations. Leave-One-Out Cross-Validation (LOO-CV) cannot be effectively used in time-series models such as these, since future datapoints depend on the existence of previous ones. Estimated LFO-CV is similar to LOO-CV but estimates the model's predictability of M future datapoints when the model is fit on only the first L observations. We ordered our datapoints by trial number and set L to equal 100 * N and $M = 5$ where N is the number of participants for that experiment. This effectively means we are assessing each model's ability to predict 5 datapoints after the removal of those 5 datapoints as well as all future ones, incrementally downwards until the first 100 trials for each participant. LFO-CV does not explicitly penalize a higher number of parameters, but a model with more parameters is not always assessed as being better, since overfitting of those parameters with the data subset would lead to poor estimates of out-of-sample observations. Estimated LFO-CV estimates a pointwise expected log predictive density (ELPD); the pointwise difference in ELPD between two models can be used to compare their relative predictive power (and a standard error of the difference). All models performed well in parameter recovery tests with simulated data.

6.7.1. Results

There was no single model that best explains all data (a). Posterior distributions show parameter fits for each model for each experiment dataset (b), and fits broadly support a gating hypothesis of credit assignment (which predicts $k < 1$) under the model assumptions. For Experiment 1, update ratio k was very far from 1 in every model tested, supporting a gating hypothesis. The shift and split gating model posterior fits for the two values of k overlapped, indicating in this experiment the gating rate did not change over time or between the two targets. This is reflected in the model fits, where these two models did were similar or even slightly worse compared to the basic gating model. However, the Miss Reward model produced a significantly better fit in comparison to the other models, and the fit revealed a miss with an equivalent reward of around 0.97 points was the most likely.

In Experiment 2, the update rate fit value for the Basic Gating model was much higher (mean = 0.74) but still clearly different from 1, implying less overall gating, but still some amount of gating. The Split Gating model makes the most conceptual sense here, as the Certainty- target should facilitate gating (no reward information) whereas the Certainty+ target should facilitate updating of target value since fictive reward information is presented. Indeed, the Split Gating model performed well with the two update ratios implying a different gating rate per target (Certainty-: 0.58, Certainty+: 0.87). However, the Miss Gating model which implies gating occurs on miss, and reward information is ignored, performed the best. The standard error of this difference is very large however, implying perhaps this pattern is very inconsistent across all of the data.

Experiment 3 involved three datasets. The three conditions presented three levels of coupling between selection and execution. Each conditions was best explained by a different model. The low coupling fit broadly showed the same pattern as that of Experiment 1, perhaps because of the similarity in probability levels. The medium coupling condition was best explained by a model that assumed different update rates for each target. This fits with the hypothesis that EPE outcomes from the EPE+ target are blamed on the target selection, rather than execution (update ratio is higher), due to the higher coupling between target selection and execution errors. This pattern is also seen in the high coupling condition in the Split Gating model fit, though the posterior distributions are slightly overlapping, indication some considerable probability that the two per-target update ratio values are the same. The high coupling condition data best fit the model that assumes gating rate shifts over time. However, this shift seems to be in the opposite direction of what would be predicted (Start: 0.53, End: 0.02), implying gating increases over time (reduced update ratio). The opposite would be predicted, because a high coupling of target selection and execution errors is hypothesised to prohibit gating. The high coupling condition produced fits (in all models) with a high inverse temperature parameter, indicating greater certainty in selection (less random behaviour). This fits with the data observations, since in this condition selections were very clearly biased towards the EPE- target.

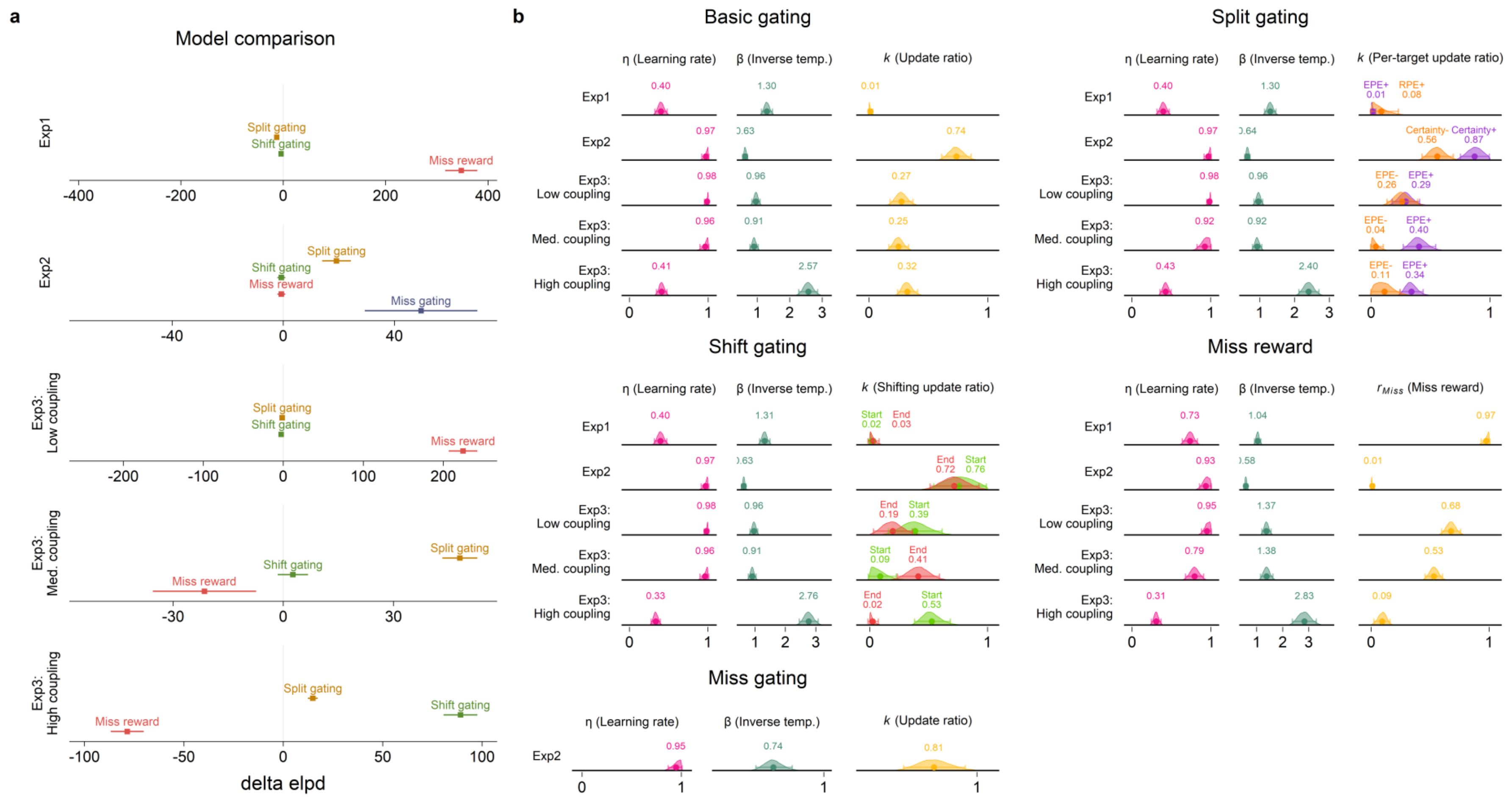


Figure 6.7 – Modelling results. (a) Model comparisons reveal the best performing model (under Leave-Future-Out Cross Validation compared to the Basic Gating model) differs for each condition. Error bars indicate standard error of the estimate of the difference (b) Posterior distributions for the five tested models, error bars indicate 95% highest density interval

6.8. Discussion

Three experiments were undertaking to examine credit assignment under selection errors, execution errors, uncertainty, and selection-execution coupling. These experiments were designed to investigate the potential “gating” phenomenon that potentially facilitates risk-seeking behaviour in a motor context, whereby participants prefer to select targets that yield more execution errors, even when average reward pay-out is identical.

6.8.1. Experiment 1: Manipulating Error Type

In Experiment 1, the apparent risk-seeking effect found by McDougle (2016), where participants showed a trend towards selection of targets that have a higher chance of resulting in execution errors, was replicated. Here, reselection rate was measured, allowing for investigation of preferences of each individual outcome. A significantly higher reselection rate was observed following an EPE outcome compared to an RPE. A gambler’s fallacy effect could be affecting selection here – a lower than expected reselection rate following a Reward could be due to the belief that it is unlikely to see a string of consecutive rewards from repeated selection (Kahneman and Tversky, 1979; Clotfelter and Cook, 1993). A tendency to reselect following an EPE could cause a selection bias towards the EPE+ target, even from the first trial as this behaviour may not be learned but instead a facet of a fundamental human behaviour.

The mechanism behind the proposed persistence effects (apparent preference to reselect following an EPE outcome) is said to be centred around error correction (Parvin et al., 2018). An RPE outcome tells us our execution was correct, but our selection was incorrect. Therefore, to fix this error we should change our selection. However, an EPE outcome tells us nothing about our selection but does inform us how to modify our execution to hit the target (ignoring effects of motor noise). The participants are instructed to always choose the target they believe is the most likely to earn them a point. With this, it follows that when we receive the EPE feedback – we have no information about our selection, all that we have is this prior that there is a higher probability that a star is contained in the selected target. Thus, the best strategy for receiving a point on the next trial following an EPE is to reselect the same target with an adjustment in execution. This holds true unless we gain belief that an execution error is in fact a property of the target (and thus not correctable) similar to an RPE.

6.8.2. Experiment 2: Manipulating Uncertainty through Fictive Outcomes

In Experiment 2, we tested the hypothesis that target selection could be driven by a desire to reduce uncertainty by selecting the target that we are most uncertain about. The EPE implicitly lacks the information about whether our selection was correct (it remains closed hiding the star or lack thereof contained within the target). Here, we fixed not only the levels of reward between the two targets, but also the rate of execution errors. Instead, we allowed one of the two targets (Certainty+) to open to reveal its value (fictive outcome) even following an execution error. We predicted that, if participants were acting on a desire to reduce uncertainty about the targets' value, we would observe a bias towards the target that hides the value from us following an execution error (Certainty-). We found there was no statistically reliable bias in selection towards either target. Thus, we found no evidence to support the claim that it is uncertainty per se that is driving the preference towards execution errors. Instead, these data provide further support for the gating account, as this model predicts no preference for either target in this experiment.

Experiment 2 also allows use to disentangle the motor component from the lack of value information normally presented following an EPE. We found that the EPE still facilitates high reselection rates, whilst the other two motor error outcomes (Fictive Reward and Fictive RPE) do not. This supports the hypothesis that RPE encodes selection information and therefore prompts switching, and that EPE encodes execution error and the lack of selection information (with our prior assumption that the selected target contains a star) means we choose to reselect with a modified execution plan. The data here indicates that when an execution error also produces selection information (Fictive Reward and Fictive RPE), participants are prompted to either reselect (on Fictive Reward) or switch (on Fictive RPE), largely ignoring the motor information. The execution is only a means to an end of acquiring points, and so it is intuitive that is gets overridden by selection information. Note that reselection ANOVAs are confounded here because they are always pitted against a different target. A more conclusive way to examine reselection behaviour independent of the differences between targets may be to perform experiments where participants must select between two targets with identical outcome probabilities.

6.8.3. Experiment 3: Manipulating Execution Error – Selection Coupling

The selection behaviour in Experiment 3 predictable favours the target that yielded the most reward on average. However, the main aim of this study was to investigate how reselection behaviour is affected by divergent levels of EPE probability. We found that diverging the probabilities of the EPE outcome between the two targets (in effect coupling execution errors with selection) elicited changed in behaviour following the RPE and EPE outcomes. Specifically, a high coupling was related to an increased reselection behaviour following RPE outcomes (across both targets – the ANOVA revealed no significant interaction), and a decreased reselection for the EPE+ target only for the EPE outcomes (there was a significant interaction in the EPE ANOVA) (Figure 6.6).

Regarding the diverging behaviour EPE reselection behaviour at high coupling, learning that a target is more difficult to hit seems to make participants less forgiving of EPEs on that target (lower chance of reselecting). Crucially, a change in difficulty of hitting the target does not seem to increase reward preference. The changing reselection behaviour for EPE outcomes implies that making a target more difficult to hit makes apparent the fact that EPEs are in fact a property of the target. As the probability of hitting either target diverges, participants may learn that EPEs are in fact dependent on the targets and believe EPEs are not correctable. When EPEs are learned to be purely a property of the targets an EPE becomes equivalent to an RPE, and therefore elicits similar reselection behaviour (reduced reselection rate). McDougle et al. (2016) found that, in a similar task to Experiment 1, patients with cerebellar impairments elicited contrasting behaviour to that seen here- avoiding a target that yielded EPE outcomes – making their behaviour most similar to that found in non-motor tasks (risk averse). Individuals with cerebellar degeneration have also been found to have difficulties with sensorimotor adaptation due to their inability to use sensory prediction error information. This implies the correctability (or belief or correctability) of EPE outcomes may be driving this apparent preference towards them and this is consistent with the data from Experiment 3 where making EPE outcomes highly likely (and thus less correctable) seems to attenuate this effect.

We found that the reselection rate following an EPE outcome was significantly higher than an RPE outcome (Figure 6.6b). A preference of reselection following an EPE outcome could drive the preference towards the EPE+ target (at least in the early stages, where no differences between the targets are learned). This implies the movement-dependent RL model cannot fully explain this behaviour, and short-term reselection behaviour should also be considered.

As in Experiment 2, the reselection behaviour in Experiment 3 is confounded because the alternative to reselection (switching) is always done towards the other target, which is not held constant here. A more robust way of testing these ideas of credit assignment might remove the free selection aspect of these tasks; force participants to select targets equally and probe their value estimation of each target directly via questionnaire (e.g McDougle, Ivry, & Taylor, 2016) or indirectly by asking them to bet points on the likelihood of a target yielding a point.

6.8.4. Modelling

Several models were tested on each dataset, but noteworthy is that no single model performed best in all cases (Figure 6.7).

The modelling for Experiment 1 revealed the update ratio to be very low across all gating models, implying the value updating following lack of reward due to an EPE was gated almost entirely. This mirrors the behavioural results, where participants selection was biased towards the target that yields more EPE outcomes, and participants were more likely to reselect following a miss. Here, the learning rate η was low and β was high, perhaps indicating a more deliberate selection process that considers many previous trials. However, the miss reward model performed the best here, with the mean value of a miss being estimated at 0.97. It is possible that missing the target might confer an advantage to the participant that is more valuable than an RPE outcome e.g. provide opportunity to improve one's motor skill which would allow the participant to maximise long-term reward.

Experiment 2 was tested under a "miss gating" model, which assumes a miss leads to gating of value updating. This differs from the basic gating model since on fictive trials; the miss gating model assumes the reward revealed on these trials is ignored. This model is estimated to perform best for Experiment 2 (though with a large standard error), which is surprising based on the reselection data, which shows clearly different behaviour based on what was contained inside following a fictive outcome. The model comparison of this model vs the Basic Gating model was very inconsistent however, with a large standard error. This implies this model is not consistently better for every datapoint. Perhaps a hierarchical model where parameters are free to vary between participants may make this result clearer.

The split gating model was designed to assess how gating might be different in EPE outcomes (Certainty- target) and Fictive outcomes (Certainty+ target). The mean fit for k for the Certainty- was significantly lower than the Certainty+ target, indicating more gating occurred when reward information was hidden. Gating was only slight on Fictive outcomes, indicating that the gating rate may be modulated not by the presence of motor errors per se, but by the lack of selection information afforded during the EPE outcome. Additionally, the update rate from the Certainty- target is much higher than in Experiment 1 which had similar outcome probabilities, suggesting that the presence of Fictive outcomes applies a global shift in credit assignment processes, raising the update rate (reduction in gating). There was also a high learning rate for Experiment 2, which suggests more chaotic selection behaviour with participant choice selection being driven almost entirely by value estimates obtained from recent trials, which is expected given the hit rates and expected values of the both targets were identical in this study. The performance of the Split Gating model in Experiment 2 fits well with the data, and the best fit update ratios are what would be expected theoretically. Specifically, less gating would be expected in the target that reveals the reward information even on miss (fictive outcome), and this is clearly observed with the larger fit update ratio in the Certainty+ target.

Experiment 3 modelling showed a distinctly different pattern of results, with different models performing best for each execution error rate condition. In the low coupling condition, where the difference between the two targets was small, the split and shift gating models predictably did not perform better than the basic model. The miss reward model performed best here like in Experiment 1. The medium coupling condition presented participants with targets that had easily detectable differences in expected value and hit rate, and this produced a best fit for the split gating model. Here, the EPE+ target update rate was found to be significantly higher than the EPE- target, supporting the hypothesis that a differing miss rate per target makes the update rate increase for EPE outcomes for that target, bringing an EPE closer to an RPE. However, this pattern does not continue in the high coupling condition, instead the best fit of the shift gating model suggests the update rate is changing over time when faced with differing hit rates or expected values. This supports the idea that these gating processes change over time when new information is learned, namely here the participants are learning that their execution errors are driven by their selection. Thus, they shift their update ratio upwards over time, therefore taking using more of the information in the credit assignment process (like they would with an RPE outcome). However, perhaps a more slightly more complex model would better fit observations in Experiment 3, combining the split and shift gating models. A different model may be able to capture the expected mechanisms of the EPE outcomes of the two targets being treated differently (as seen in the reselection results), where that difference would emerge over time as it is learned. The shift gating model is confounded by selection

The miss reward model fits some of the datasets well (Experiment 1, Experiment 3: Low coupling), and so future studies could examine this potential element hypothetical reward on a miss, by presenting participants with a choice between monetary reward and implicit reward given on misses. It is also possible that this observation may be the result of a high reselection rate following misses, which none of the models capture. This then begs the question if any of these observations are dependent on a "miss" outcome being present. Unless manipulated to be such, it seems here that an execution error is seen as a fully intrinsic error, i.e., all blame for that outcome lies on the participant's actions, not on the target they selected. It may be possible that any task that produces errors that can be believably blamed on the participant may facilitate this type of behaviour, even if that task is non-motor.

6.9. Conclusion

Here, we have replicated the previous behavioural observations of tendency to select a target that yields more execution prediction errors over reward prediction errors in Experiment 1. Next, in Experiment 2 we showed that this behaviour is not driven by the lack of reward information afforded following incorrect execution. In Experiment 3 we demonstrated that the gating process could be modulated by increased coupling between selection and execution. Finally, a series of models were constructed, fit to the data, and compared in their quality of fit for each experiment. Three datasets (Exp 2, Exp 3: Medium, High coupling) were best explained by modifications to the movement-dependent reinforcement learning model first proposed by McDougle et al. (2016), while the other two (Exp 1, Exp 3: Low coupling) were best explained by a model where misses are assumed to be rewarding. The data here support the claims of a gating mechanism in credit assignment, and go further to show how the mechanism can be altered with presentation of reward information, and coupling of selection and execution. More broadly, this study shows the complex role of error in motor learning, and how the interpretation of the error by the system influences learning. Future studies should aim to manipulate the role of the motor component of the task to better understand human credit assignment processes.

7. Does the Gating of Reinforcement Learning Generalise to the Cognitive Domain?

7.1. Overview

Decisions are implemented by actions, and when an action fails to provide the expected reward, a credit assignment problem must be resolved: Was the error due to a poor decision, or just a poor execution of the movement that implements the decision? Recent studies have concluded that credit for a decision's lack of reward offering can be discounted when there was an error in movement, thus affecting the perceived value of the selection. Current interpretations suggest this is the product of a unique interaction between movement and decisions. However, the limits of this phenomenon are not known. This study investigates whether this potential mechanism driving value estimation processes might similarly manifest in a non-motor task. To this end, we created a multi-stage cognitive decision-making experiment that was an analogue of the action bandit task described in the previous chapter. Specifically, a two-alternative forced choice task was developed where participants selected one of two treasure chests to obtain rewards. To access the chest (and potentially obtain a reward), a key must first be found, apparently hidden inside one of two cups. The probabilities of one chest were set such that participants had a higher chance of failing to find a key, the other a higher chance of finding a coin once a key was obtained, but both ultimately offered equal chance of reward. Observation of selection behaviour showed a preference for the target that facilitated more errors in acquiring the key. The behavioural pattern was very similar to that in the previous action-bandit studies and is best explained by the participants’ false priors of the task structure, as they are led to believe selection had no impact on the chance of finding a key. Parameter fits of a mechanistic model of these processes support this claim, showing the data can be explained by assuming selection does not affect the chance of finding a key. These results propose a new way of looking at the broader credit assignment systems in human learning, implying a model-based understanding of a system can influence model-free learning processes.

7.2. Introduction

The process of making a decision entails first deciding which action plan to take (i.e. selection) and then implementing the plan by interacting with the world (through motor execution). The majority of research on decision sciences (i.e. in the fields of behavioural economics and experimental psychology) have historically ignored the role of implementing a selection in modelling decision making (Taylor et al., 2014; Galea et al., 2015; Aczel et al., 2018). This is apparent, for example, with experimental designs employed by researchers in these domains – with decision execution being operationalised via rather trivial processes, e.g. button press on a stimulus response pad or through verbal response. Yet, the process of executing a decision in the real world, beyond laboratory studies and survey methodologies, often brings with it much greater sensorimotor demands. Think for example of a surgeon, mid operation, deciding on the next incision to make. Her choice will invariably be driven by the degree of difficulty in being able to physically execute the incision in a smooth and efficient manner. These ideas have long been promoted by embodied theorists who propose a bilateral relationship between action and cognition (Wilson, 2002). For the process of decision making specifically, there is now growing evidence that these selections can be modulated by execution demands in a multitude of ways (Green et al., 2010) but the mechanisms underlying this relationship remain unclear.

One specific topic we have grappled with in the preceding chapters is how the process of reinforcement learning proceeds in the presence of an error in execution error. A recent motor dependent reinforcement learning account postulates that execution errors gate the process of value updating associated with an action plan (McDougle, Boggess, et al., 2016; Parvin et al., 2018; McDougle et al., 2019). In this way, poorly implemented actions do not (largely) change estimates of appropriately selected actions.

Behaviourally, experiments designed to test this idea, by introducing options that have averting degrees of execution and selection error have found a consistent bias towards the selection of a target that yielded more execution errors. One potential explanation, that the bias towards these options is driven by a need to reduce uncertainty about value estimation (because information about the appropriateness of the selected strategy remains unknown when non-rewards arise from execution errors) was ruled out through an experiment reported in the previous chapter. Briefly, this experiment forced the participants to select from two targets, both targets had an equal chance of hit or miss and equal chance of reward following a miss. The only difference between the two targets is that one target revealed what the outcome *would have been* given the participant were to correctly execute their selected action, resolving the uncertainty that is normally present on an execution error. Under the hypothesis in question, participants should select the target that facilitates uncertainty, through hiding what would have been awarded on a successful action. In fact, participants showed no significant selection bias, which is supported by the gating hypothesis.

A subsequent experiment obtained evidence that this gating rate was malleable, i.e. coupling of execution errors with target selection through differences in hit rate between targets would reduce gating to some extent, and also perhaps the gating rate is different between selection.

In previous formulations, the experimental designs have deliberately parsed selection, execution and outcome so that they are dissociated. In the first stage, the participant selects a target they believe allows them to maximise their points. In the second stage, the participant attempts to interact with the selected target by swiping their arm to try to collide a cursor with it. In this way, the two stages are seemingly dissociated, i.e., the chance of hitting is not related to the selected target. Could it be that this dissociation is the driving factor in choice perseveration?

Parvin, et al. (2018) showed that the selection bias effect towards the target that yields more execution errors is influenced not by the strength of the sensorimotor prediction error, but instead by the sense of control (agency) that participant had during the task. In this way, an execution error provides a very salient correctability signal (indicating the direction and magnitude of the subsequent correction needed to be successful), therefore manipulating the participants’ belief of their role in the outcomes. In effect, those that were told they were in control of the hit/miss outcome believed that the motor task at hand was a feature of themselves (not the task), and so was unrelated to the chosen target. This led to execution error outcomes updating the value estimate of the chosen to a lesser extent, compared to those who had their sense of agency reduced. Those with their sense of agency reduced could have associated execution errors with the chosen target (as they believed they were not in control); thus, the target value estimate was updated. A side effect of a greater sense of agency is a belief that the second stage (i.e. the reaching movement) is independent of the first stage (target selection).

The present study extends these ideas and tests the hypothesis that it is not the agency that drives this behaviour per se, but the behaviour is instead driven by the general internal representation of the task. Here, we create a two-stage decision-making that removes sensorimotor demands from the second stage. In this two-alternative forced choice, participant must select the treasure chest which they believe is most likely to contain a coin. However, after selection, participants must guess which of two cups contains a key that is required to open the chest. Importantly, the cup guessing game is presented as being independent of the selection made in the first stage and there is no opportunity to correct or accurately implement the selection at this stage of the process.

We hypothesise that any errors made during the guessing game will not be attributed to the target directly (consistent with the credit assignment gating hypothesis), but instead (to some extent) be fully attributed to the guessing game itself. Evidence for a gating mechanism is expected in the form of a preference for the chest that facilitates more errors in the cup guessing stage and a difference in reselection rate following different error types. If this relationship holds, it will provide a demonstration of how an internal model of the task, given through instructions as well as the visual representation of the task, can modulate the credit assignment process, and cause potentially sub-optimal behaviours independent of the sensorimotor system. More broadly, this will bring into question the true nature of cognitive-motor interactions

7.3. Methods

7.3.1. Treasure chest task

This experimental task mimics the task structure of the previous action-bandit task (Chapter 6), but with a thematic change, replacing the motor component (arm reaching) with a guessing task. On each trial, participants were presented with the choice of two “treasure chests”, which they initially selected through clicking on it with a mouse. Then, the participant took part in a separate game, in which they must try to find a “key” hidden under one of two cups (participants were told there was a 50% chance of finding the key). If they managed to find they key, the chest they chose opened to reveal a coin (1 point) or nothing (0 points). The cycle then continued, the participant once again must make a choice between the same two chests, and were told verbally and with on-screen instructions to select the chest they believe is most likely to lead to a reward. Directly comparing this to the VR action-bandits task, the chest selection replaces the target selection, and the cup selection replaces the action execution (Figure 7.1).

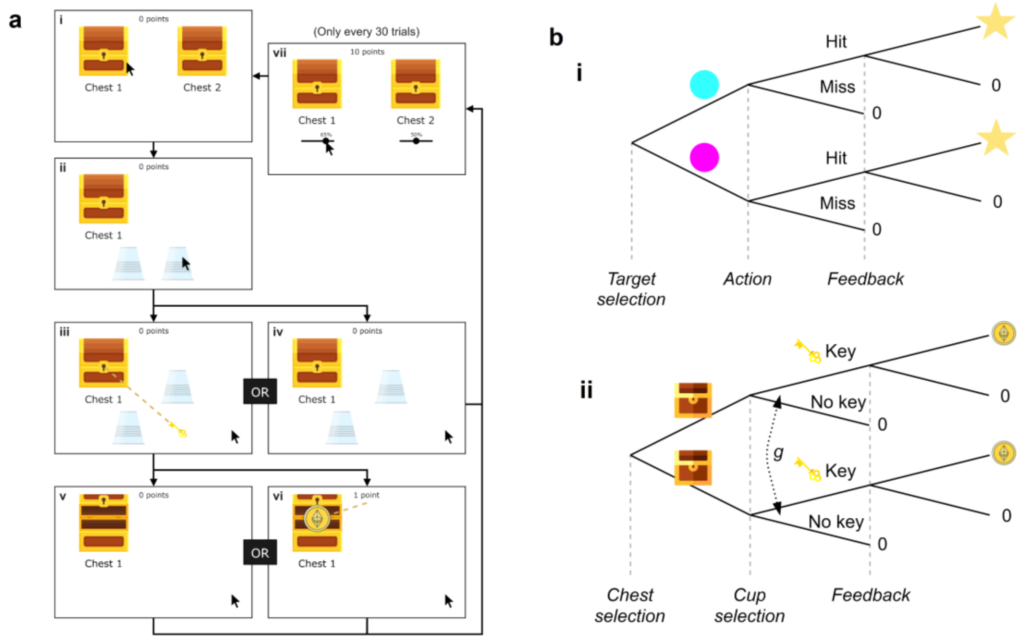


Figure 7.1 – Treasure chest task design. (a) (i) The task began with a selection of a chest with a mouse. (ii) Participants guessed the location of a key contained within the cups. (iii) The selected cup may reveal a key, or (iv) may be empty and reveal nothing, ending the trial with 0 points. (v) If the participant did find a key, the key unlocks the chest, revealing either an empty chest for 0 points, or (vi) a coin, worth 1 point, which is immediately added to the participant's score. (vii) Every 30 trials, a "probe" takes place before the chest selection. It required the participant to input their belief, in percentage terms, of them finding a coin following hypothetical selection of both chests. (b) Comparison of two-stage tasks, (i) the original action-bandit VR task used in previous studies, and (ii) the current treasure chest task design. Both tasks crucially are made up of two seemingly independent stages. Failure to complete the second stage (action execution/cup selection) prevented participants from receiving feedback from the initial selection (target/chest). *g* represents the generalization between the cup selection following either chest selection (see modelling section).

The experiment was performed on a 1920x1080 24" monitor, with participants interacting with the task using a mouse. The task was developed using the Unity Experiment Framework (Brookes et al., 2019). The experiment consisted of 300 trials. Every 30 trials a probe trial took place. Before chest selection, participants used the slider to specify their belief of the likelihood of receiving a coin if they were to hypothetically choose that chest at this point in time (in percentage terms; Figure 7.1a.vii). The sliders were set at 50% for the first probe trial, but then remained at whatever value the participants specified for the following probe trials (i.e. they were not reset). The sliders displayed the numerical percentage value (rounded to the nearest 5%). Participants were forced to click (but not necessarily move) both sliders before beginning the trial to ensure deliberate input of percentage estimates.

7.3.2. Outcomes

The two-stage task (Figure 7.1c) ensured that there were 3 possible outcomes. The only way a participant could earn a point is to succeed in both stages; i.e. find a key, and the chest opens to reveal a coin. This was the *Reward* outcome. If the participant failed to find a key during cup selection, this was a *Cup error*. If a key is found but opened to reveal an empty chest, this is a *Chest error*. These mimicked the two error types crucial to the previous bandit tasks, with a Cup Error analogous to an execution error, and a Chest Error analogous to a selection error.

The probabilities of these 3 outcomes had their probabilities fixed, with the outcome of each trial being predetermined. The probabilities were set differently for each chest, manipulating the distribution of the two error types. Chest 1, CupError+ had a higher probability of yielding a cup error, and ChestError+ had a higher probability of yielding a chest error (Table 8). These two chests directly mapped to the EPE+ and RPE+ targets, respectively, in Experiment 1 in Chapter 6. Crucially, the chance of reward was identical, controlling for any preference towards rewarding outcomes. The position of the two chests was counterbalanced between participants such that any positional preferences washed out.

Table 8 – Outcome probabilities for the two chests. The two chests differed only in their distribution of the two error types and had an identical expected value.

	CupError+ chest	*ChestError+ chest*
Reward	30%	30%
Chest error	20%	50%
Cup error	50%	20%

7.3.3. Participants

18 participants were from the University of Leeds participant pool scheme, each being paid £5 on completion of the task, with a session lasting around 20 minutes. The Psychology Research Ethics Committee at the University of Leeds approved the research.

7.3.4. Statistical analyses

We measured target selection bias, i.e. the percentage point (p.p.) difference between the overall selections of the two chests. A one-sample t-test was performed comparing the participant's overall selection biases to 0.

Reselection preference is a measure of how likely a participant was to reselect the same chest following each outcome but controlled to an individuals' overall reselection rate. This was calculated by first computing the percentage of trials where the participant reselected the same target following each outcome for each chest and subtracting from this the percentage of trials where the participant reselected the same target (regardless of outcome). Here, a within-subjects ANOVA was performed, with reselection preference as the outcome variable, and the previous trial outcome, previous trial chest, and the interaction between these two were used as independent predictors. Cohen's d effect sizes are reported where appropriate. All data met assumptions of normality through assessment by histogram, Q-Q plots, and Shapiro-Wilk tests.

The mean estimate bias was calculated, i.e. the percentage point difference in reported value estimates from the probe trials, averaged across the session, to assess if the reported value bias was significantly different from 0. This was used rather than the final probs as the probe values were likely subject to recency effects, and therefore a single probe report may present a more noisy measure. Additionally, this per-participant mean estimate bias was regressed against their overall selection bias. This was to confirm an individual's behaviour reflected their estimates of chance of obtaining a coin (as per the task instructions).

7.3.5. Model

The hypothesis involves the generalisation of learned value between different parts of the task. Specifically, it is assumed that participants are estimating the probability of completing both stages of this two-stage experiment individually – however, when updating probability of completing the cup guessing game, the value updating process is not fully associated with the initial selection they made, but instead generalises across to affect the value estimate of the cup game within other state (Figure 7.1b.ii). This builds on the gating model, but simplifies and generalises it to accommodate any two-stage task such as these.

The model utilises a basic model-free learning formulation i.e. the form initially proposed by Rescorla and Wagner (1972). However, in contrast to the previous work, here we assume value learning of each of the two steps independently. In this model, when we select a cup, we update our estimate of probability of the cup providing a key based on the difference between the current probability estimate of finding a key after selecting chest x ($p_{k,t}(x)$), and the presence of a key on that trial (k_t). We add that difference to our current probability estimate for that chest with a learning rate, η,

$$p_{k,t}(x) = p_{k,t-1}(x) + \eta\left(k_t - p_{k,t-1}(x)\right). \qquad 19$$

Crucially, we also perform this operation with cup stage on the non-chosen chest, y, and multiply by a generalisation rate g. This simulates the generalisation of the learning that occurs when estimating the probability of finding a key after selecting chest x to the chest y.

$$p_{k,t}(y) = p_{k,t-1}(y) + g\eta\left(k_t - p_{k,t-1}(y)\right). \qquad 20$$

The hypothesis here is that this parameter g is non-zero. Since participants are told the chance of finding a key is a constant 50%, and is unrelated to chosen chest, participants should begin the task with a strong prior that g in fact equals

1 (in effect updating the key probability estimate related to both chests concurrently).

If the participant finds a key, they can learn about the contents of the chest they initially selected. This is modelled in the same way as above, by assuming participants are learning a probability of finding a coin within the chosen chest $(p_{c,t}(x))$, and updating their estimate using the observation of presence of a coin on that trial (c_t)

$$p_{c,t}(x) = p_{c,t-1}(x) + \eta\left(c_t - p_{c,t-1}(x)\right). \qquad 21$$

However here, there is no generalisation assumed (participants have no real reason to believe the probability of finding a coin in one chest affects the probability in the other). Therefore, the probability estimate of finding a coin is simply carried over for the non-selected chest (y).

$$p_{c,t}(y) = p_{c,t-1}(y). \qquad 22$$

The combined probability estimate of receiving a reward on a given trial is therefore the compound probability of both finding a key and finding a coin (both are required for a reward):

$$p_{r,t} = p_{k,t}p_{c,t} \tag{23}$$

This reward probability estimate can then be directly compared to the probability estimates the participants provided during probe trials (for both chests). As probability estimates were only collected every 30 trials, these values were linearly interpolated between these probe trials. To link the model to the probed estimates (and account for inevitable variability), the probed estimate was assumed to come from normal distribution on each trial for each chest, with the model estimate being the mean of the distribution and an unknown parameter σ being the standard deviation.

$$\mathrm{probe}_{r,t} \sim \mathrm{normal}\big(p_{r,t}(x), \sigma\big) \tag{24}$$

Bayesian inference was used to estimate the values of g, η, and σ. We estimated the posterior distribution $P(\mathrm{Parameters}|\mathrm{Data})$ using the No-U-Turn algorithm (Hoffman and Gelman, 2011) implemented RStan 2.19.1. The Gelman Rubin statistic ($\hat{R}$) (Gelman et al., 2014) which assesses convergence was equal to 1 for all parameters. The values of η and σ were fit per participant, with g fit across the dataset. Eight chains of 4,000 samples were taken from the posterior; the first half were discarded as warmup samples. There were no probe values for the first 29 trials, so these trials were not used in the inference (but probability estimates were still computed for these trials). Both g and η were assigned uniform distribution priors between 0 and 1, and a uniform prior between 0 and 100 for the σ parameter. The initial probability estimates in all cases were assumed to be 0.5 on the first trial. This model performed well in parameter recovery assessments, where the model was fit to data that were simulated using fake parameter values.

Code for generating the figures, performing the analysis, and performing the modelling are openly available at https://github.com/jackbrookes/treasure-R-project.

7.4. Results

The selection bias was found to be significantly biased towards the CupError+ chest (Figure 7.2a); *mean* = +12.6 p.p. bias towards CupError+, *t*(17) = 2.19, *p* = .043, *d* = 0.516. This indicates a preference for the chest that facilitated more Cup Error outcomes over Chest Error outcomes, mirroring the effect found in the VR action bandit experiments.

The reselection preference analysis revealed a significant main effect of the previous trial outcome on the reselection rate; *f*(2, 17) = 7.69, *p* < .001, η^2_p = 0.147. The main effect of chest selected was not statistically significant; *f*(1, 17) = 0.95, *p* = .333, η^2_p = 0.010 and there was no significant interaction; *f*(2, 17) = 0.602, *p* = .550, η^2_p = 0.013 (Figure 7.2b). Pairwise comparisons of the mean reselection rate across three outcomes (collapsed across target) revealed reselection rate to be significantly different across the two error types (Cup error vs Chest error); *p* < .001; the other two pairs were not significant (Cup error vs Reward, *p* = .131; Chest error vs Reward *p* = .067).

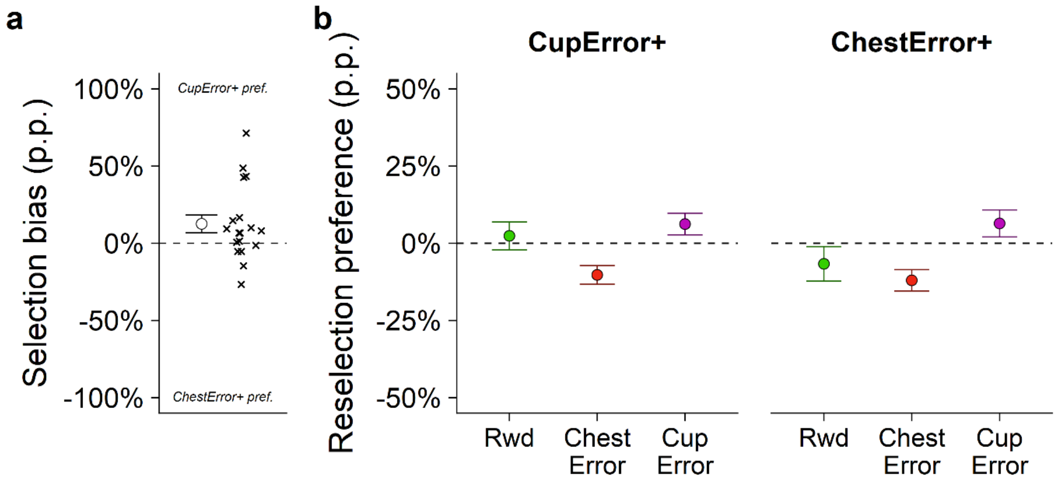


Figure 7.2 – Behavioural results (a) There was a statistically significant selection bias with a preference towards the CupError+ chest. (b) There were statistically reliable differences in reselection preference across trial outcomes. This was driven by a higher preference for reselection following Cup Error outcomes. Error bars show S.E.M.

The mean estimate bias (average difference in reported estimates on probe trials) was not significantly biased either way (Figure 7.3a); *mean* = +12.6 p.p. bias towards CupError+, *t*(17) = 0.99, *p* = .336, *d* = 0.233. However, the linear regression of an individual's mean value estimate bias vs their selection bias was found to be significant, slope = 1.29, R^2 = 0.65, *p* < .001 (Figure 7.3b). This indicates an alignment between selection and value estimates, providing evidence that participants selections were driven by high level value estimates.

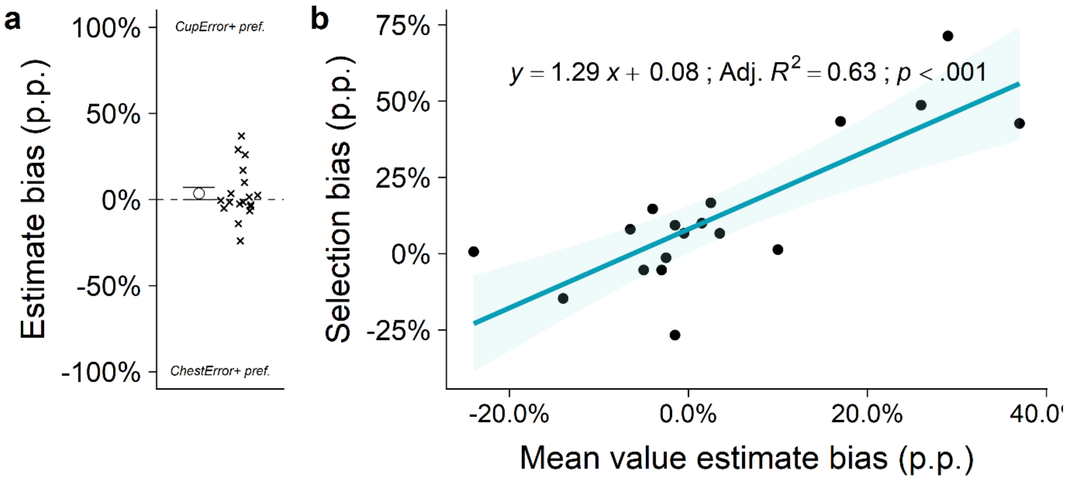


Figure 7.3 – Reported value estimates (a) The mean estimate bias (average difference in reported estimates on probe trials) was not found to be significantly biased towards either chest. Error bar indicates S.E.M. (b) There was a significant positive correlation between the mean estimate bias and the selection bias. Solid line shows best fit with shaded area +/- 1 S.E.M of the slope. Values expressed as percentage point differences (p.p.).

The model fits support generalisation across the two targets in the cup stage, as the fit indicated the data supports a mean value of $g = 0.97$ (Figure 7.4).

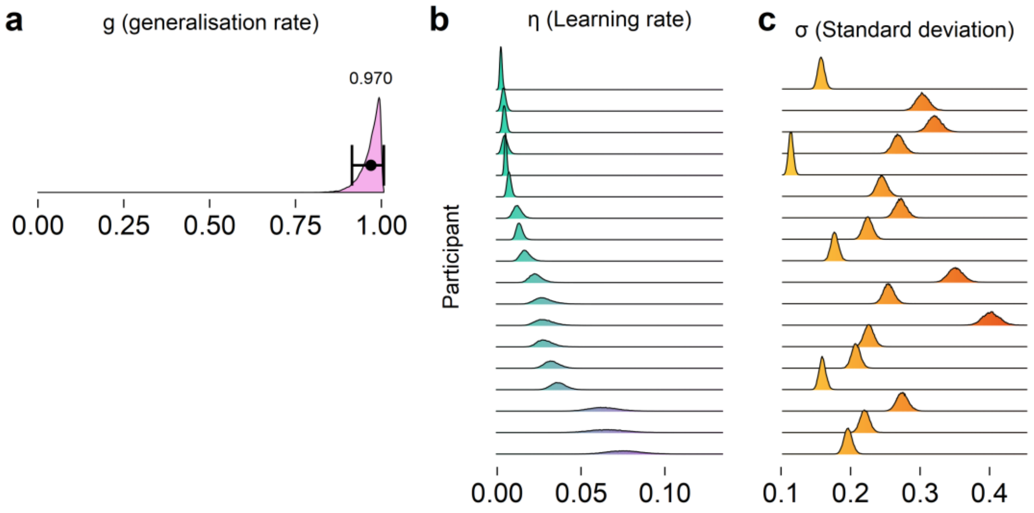


Figure 7.4 – Model fits (posterior distributions) for generalization model using STAN. (a) The generalization rate fit revealed most participants' data was best explained by a generalisation rate close to 1. Point and text indicate mean of samples, and error bars indicate 95% density interval. (b) The learning rate parameter here per-participant, ordered from smallest to largest mean fit value. (c) The standard deviation of the model residuals (assuming they are normally distributed) was fit per participant.

7.4.1. Discussion

This study investigated the effects of an intermediate, seemingly independent, stage in a bandit task on decision making behaviour. Previous studies (such as the action bandit task reported in the previous chapter) operationalise this intermediate stage as a motor task (e.g. hitting a small target), which must be completed before revealing the reward (or the absence of reward) afforded by the chosen bandit. It has been assumed that behaviour that arises from these experiments (selection biases, differing reselection rates between error types) are a feature of a cognitive – motor interaction mechanisms.

To account for such phenomenon, one hypothesis was that the motor system interjects during the credit assignment process and partially prevents (or gates) the value updating (McDougle, Boggess, et al., 2016). Parvin et al. (2018) showed how this effect is facilitated by agency in the motor stage of the task, rather than the strength of the execution error feedback. Here, we find this pattern persists even when the intermediate stage is non-motor and even non-correctable. We discuss the implications of this finding below.

The results of the selection bias t-test show a significant bias towards the CupError+ target (Figure 7.2a). This mirrors the direction of the effect found in the previous studies, where there was a bias towards the target that yielded more execution errors. A similar trend of behaviour is seen in the reselection preferences (Figure 7.2b). A significant difference can be seen across outcomes, and in particular between the two error types (CupError vs

ChestError). This was also found in previous studies, where the execution prediction error outcome facilitated significantly higher reselection rates compared to the reward prediction error outcome. This was originally seen as potential evidence for the gating hypothesis- it was clear that people did not feel inclined to switch away following an execution error, which implied that the outcome had not reduced their estimate of the value.

One importance difference in this experiment was the inclusion of probe trials. Every 30 trials participants were forced to express, in percentage terms, the chance they believe they have of finding a coin following hypothetical selection of both chests. This allowed for measurement of the high-level (explicit) value estimates held by the participants. This was an important addition given the modelling approach employed, since two absolute probability estimates are modelled in the generalisation model; selection behaviour only reveals relative estimates. Surprisingly, the mean estimate bias reported on the probe trials were not found to be significantly different from 0. This could be due to a lack of power, or maybe that the value estimate effects investigated here are on an implicit level, and participants do not hold an explicit bias in the target value. This idea may be investigated further with an experiment that measures explicit value estimates more subtly e.g. perhaps by asking participants to take bets on how likely they are to find a coin inside either chest. Encouragingly, the estimate biases were significantly positively correlated with participants' selection biases, i.e. participants who expressed a preference for one chest over another (through the probe trials) were more likely to select their preferred chest more often (in line with the given instructions).

The generalisation model mathematically articulates the hypothesis tested here. It states that participants understand the principles of the task (i.e. that one must gain the key and coin to earn a point) but learn the probabilities of passing each stage in a model-free manner. It assumes a generalisation rate, g, transfers some of the learning with regards to the cup stage within the context of one chest to the other chest. The amount by which they transfer this learning indicates whether they believe the chance of finding a key is unrelated to the chosen chest. If participants believe they are unrelated (i.e. $g = 1$, all learning is transferred) then there is no surprise that participants seem to value the CupError+ target more than the ChestError+ target. In this case, all that matters is the probability of finding a coin from either chest given a key is found; CupError+ offers 30 / (30 + 20) = 60%, ChestError+ offers 30 / (30 + 50) = 37.5% (calculated from figures in Table 8). Previous models do not include the participants' estimates of their chance of hit or miss (in this experiment, finding a key). This model better reflects the two-stage structure of the task, i.e. a bandit task with intermediate stage which must be completed before revealing the contents of the bandit.

The Bayesian inference (Figure 7.4) produced a posterior distribution which showed the value of g to be close to a value of 1, which supports this hypothesis. The learning rate was found to be much less for most participants than previous studies, but here the model is fit based on probe value estimates, which were only collected on a subset of trials. These values were linearly interpolated between probe trials, and so a small learning rate would be expected in this case. It is clear however that the model does not predict the reported value estimates very accurately due to the high σ values it produced as best fits. If value estimates were reported on each trial this might be expected to be more accurate.

The fact the behavioural data mirror that shown in the previous study, and the model fit here produces a value of g being close to 1, is strong evidence that credit assignment processes are driven by the internal model of the task structure, rather than a blind value estimate for each option. The influence of models on choice behaviour has been previously noted in literature (Daw et al., 2011). Moran et al. (2019) came to a similar conclusion following a task that presented participants with retrospective information about the task that they could use to preform credit assignment. They pose this as model-based inference, through resolving uncertainty, guiding a model-based learning process. This present study seems to be also be an example of this effect, as when the intermediate cup game stage is presented as independent of chest selection, selection biases occur.

This idea could be tested more explicitly in future studies, with a secondary condition that gives participants a strong prior of $g = 0$. This may be executed via task instructions (e.g. telling participants the chance of gaining a key is related to the chest they initially select), and/or changes in the presentation of the task. For example, if the cups were positioned near the chosen chest, it would have been clear that this cup guessing game was different to the one that would have been shown in the other chest.

7.4.2. Conclusion

This experiment aimed to investigate whether the processes of credit assignment during multi-stage tasks are driven by an assumed model of the task. Previous work has found support for a gating mechanism, whereby the credit assignment process is attenuated when an execution error is made. Here, an amended hypothesis is put forward that broadens the predictions of the hypothesis. It states that the behaviours observed in these tasks such as selection biases are driven in-fact by an (incorrect) set of assumptions about the task. Here, a task is set up that provides the illusion that the intermediate stage (cup guessing game) is unrelated to the initial chest selection. In the motor equivalent tasks, participants presumably enter with a strong prior that both targets are equally easy to hit. Here, an almost identical pattern of behavioural data is observed, even with the replacement of the motor aspect of the task, providing strong evidence that these effects do not arise from a unique interaction between motor and cognitive processes.

8. General discussion

8.1. Overview

In the following sections we reflect on the key ground made in the preceding chapters and explore the implications of this work for theory and application and offer avenues for future research.

In brief, this thesis has introduced novel tools and tasks to probe the processes underlying sensorimotor learning and decision-making. Specifically, the experimental work has built on, and refined, existing models of sensorimotor learning and decision-making from the literature and tested these models empirically. The thesis has also introduced a new research tool and it is hoped that longer term this software will allow future researchers to take advantage of VR technology- technology that has the potential to fundamentally transform the study of human learning and decision-making.

8.2. An Information-Theoretic Account of Learning

Chapter 2 began with a re-analysis of data collected from previous motor learning studies and tested the idea that an information-theoretic account could resolve the apparent paradox, where experiments on training under force fields with disruptive forces can accelerate learning (Sigrist et al., 2013). While several experimental paradigms have observed these learning processes, there is still a need to more deeply understand the mechanisms that govern them (Heuer and Lüttgen, 2015). We reasoned that one potential explanation of this phenomenon may be that disruptive forces drive exploration of the parameter space of the task, thus facilitating the development of internal models that allow for making better predictions about the task (Wolpert et al., 2011). Specifically, in Experiment 1, error-augmenting disruptive forces, assistive forces, or no forces, were used in to manipulate the movement of the participant as they used their arm to track a moving target. This moving target took place in a static workspace, which exhibited static position-dependent forces (in addition to the error augmenting forces) on the participant's hand. In line with previous work, disruptive forces facilitated learning. In Experiment 2, the error augmenting forces were tuned to performance or randomly selected on each trial. Here, the randomly selected force profiles facilitated the best learning. Re-analysis of the data from these experiments were performed using a model which quantified the amount of workspace exploration the participant had performed throughout their training. Crucially, exploration here meant that participants were exposed to the forces that underlie the workspace.

Skilled performance in this task relies upon participants either: making accurate predictions about the forces experienced across the workspace (thus are able to prepare subsequent counter forces), or refining their ability to counteract the workspace forces on-line as they are experienced. Hypothetically, these would be enhanced through greater workspace exploration, and thus participants with greater workspace exploration should show enhanced learning. Application of this model to these data showed that workspace exploration was able to explain 12% of the variance in the learning in Experiments 1A and 1B. These results align with previous research, whereby amplification of the dynamics of a task facilitates learning (Emken and Reinkensmeyer, 2005). However, these data did not show cause and effect conclusively, as the intervention forces were different between groups, thus a direct causal link cannot be inferred.

In Chapter 3, information acquisition was more directly manipulated. The path trajectory during training was designed such that it either promoted more exploration of the workspace (High Variability condition) or less exploration (Low Variability condition). The high exploration condition was exposed to more workspace information according to the proposed information model, however this did not improve learning outcomes. Therefore, exposure to the workspace here was not directly causing an increase in learning. Returning to Chapter 2 where disruptive forces were found to enhance learning, it is clear that there must be some other feature of disruptive forces, aside from the enhancement in workspace exploration, that benefits learning outcomes. Perhaps, given the crucial role for errors in learning (Laura Marchal-Crespo et al., 2014), the beneficial effects of disruptive forces that facilitate learning manifest only when interacting with error-based learning mechanisms.

8.3. A New Tool to Investigate Motor Learning and Decision-Making

In Chapter 4, the development of the Unity Experiment Framework (UXF) was reported. The methodology of previous chapters highlighted the need to improve the development of experiments for motor learning. In recent years, new advancements in hardware and software have made it feasible to use Virtual Reality (VR) technology to study human behaviour and these systems are ideally suited to the examination of the sensorimotor system. For example, complete control of the visual stream allows convincing manipulation of feedback such as errors. However, using these technologies often requires technical skills beyond the expected capabilities of a typical behavioural scientist. UXF alleviates some of the technical burden placed upon researchers, by forming a common experimental model that the researcher builds upon. This experimental model implements many features that are laborious to develop, but provide useful functionality for running experiments. For example, UXF automatically handles formatting and saving data files for trial responses or movements with appropriate headers, file names, and predictable directory structure. Since creation, UXF has been used to create dozens of tasks and has facilitated the data collection of hundreds of participants. UXF was used to create tasks for all experiments in subsequent chapters.

8.4. Capturing Redundancy and Bliss Across the Motor Learning Process

In Chapters 2 & 3, the question of the mechanisms underlying motor learning were investigated. Recall, in the role of information acquisition was investigated, after evidence was found that motor learning was enhanced when participants were exposed to disruptive forces, which facilitated more errors as well as exploration of the task workspace. This led to the conclusion that perhaps the workspace exposure was not the driver of motor learning; instead the role of the increased amount and frequency of errors are likely crucial (Milanese et al., 2008; Laura Marchal-Crespo et al., 2014). In this study, the role of this enhanced error signal was investigated, but here made independent of a haptic intervention. Specifically, error was visually enhanced by offsetting the participants cursor by some amount when their cursor was close to the target. This intervention was tested verses a no intervention control group, and performance was assessed without the intervention in both cases after a 2-week training period. Ultimately, no evidence was found that the error amplification intervention enhanced motor learning. This helps the field progress forward because it provides further evidence that the presence of the force itself in a disruptive force intervention plays a key role in the accelerating learning. There is a clear need to understand the mechanisms of enhanced learning interventions (Heuer and Lüttgen, 2015), as motor skills are vital for everyday life, and outstanding motor skills are demanded in certain professions.

The motor learning task platform that this experiment is built upon allows for investigations of other fundamental aspects of learning. The task was designed to have explicit redundancy, as four hand rotations were used to move a 2D cursor towards a target. Motor redundancy is an important feature to understand in human movement, first noted by Bernshteĭn (1967) In a redundant task there was a range of valid solutions for any given target, lying on a “manifold” (parameter subspace) (Scholz and Schöner, 1999). That means solutions can be examined between trials, and a measure for the “null-space” variability between trials can be calculated (Latash et al., 2001; Scholz and Schöner, 2014). These analyses were used to find corelative relationships between this null-space variability and the learning outcomes. Specifically, analyses confirmed that (as designed) null-space variability is not directly indicative of performance, and is as expected independent (Scholz and Schöner, 1999). However, it was found that there was a relationship between the rate of reduction of null-space variability and performance improvement, implying perhaps some advantage is gained in the ability to perform consistent solutions. No evidence was found for null space variability being predictive of subsequent learning, as was observed with task-space variability by Wu et al., (2014) Future studies should aim to further investigate the causal link between these aspects of learning (Cardis et al., 2017). Furthermore, in order to better assess the effects of solution variability, future studies could restrict the redundancy to examine if those with small null-space variability were still able to perform better when their preferred solutions were no longer available.

8.5. Resolving the Credit Assignment Problem: Motor-Dependent Reinforcement Learning?

In Chapter 6, we investigated the nature of motor errors in higher order choice selection tasks. Specifically, we tackled the question of how the brain updates value when an action fails to produce an expected reward. Should we blame the way the action was executed, or was it the fact that we selected the wrong option in the first place? This question plays out many times in everyday life, as the choices we make are implemented by physical movements. A more general formulation of this question is often referred to as the "credit assignment problem", which concerns resolving the ambiguity in the sources of rewards or errors (Minsky, 1961; Sutton, 1984; Fu and Anderson, 2008; Stolyarova, 2018). The experiments performed here examine the effects of the apparent resolution of the source of errors, specifically when a lack of reward is observed following a movement (McDougle, Boggess, et al., 2016).

It is well established that people act in a risk-averse manner in decision making tasks with uncertain outcomes (Kahneman and Tversky, 1979). Recent studies have shown a remarkable reversal of this behaviour when decisions are implemented by movements, and a lack of rewards can be interpreted to be caused by movement errors (McDougle, Boggess, et al., 2016). The proposed mechanism that explains this behaviour is that of a gating, or attenuation, of the learning process (i.e. value updating) when an error in execution is made.

The first experiment in Chapter 6 replicates this behaviour in a new virtual reality task, with the probability of encountering execution errors (miss) or selection errors (hit, but no reward) differing between two targets. Here, a preference for the target that yielded more execution errors was observed, despite both options producing the same amount of reward. The gating hypothesis predicts this behaviour since the value estimate reduction that occurs on execution errors would be attenuated, thus inflating its perceived value.

These results however could potentially also be explained by an alternative hypothesis – that this behaviour is the result of a desire to reduce uncertainty e.g. Cohen et al. (2007). A key feature of an execution error (in contrast to a selection error) is that it does not provide any information on the value of the selection. As such, there is greater uncertainty about the value of a target that is more difficult to hit (more execution errors). Perhaps the bias towards selecting this target is due to a desire to reduce this uncertainty, rather than a perception of its high value.

To test this idea, participants were tasked to select between targets that did not differ in their distribution of errors, but only in if they offered information on what the target *would have* given (reward or nothing) if the participant were to hit. This removed the uncertainty that is normally present on an execution error on that target. If this hypothesis were correct, a bias towards that target that did not reveal the potential reward information on miss should be observed. The data revealed no evidence for this, therefore supporting the gating hypothesis over an uncertainty reduction hypothesis.

The third experiment focused on the flexibility of the gating mechanism. Here, the two targets were presented with probabilities that meant that the chance of an execution error was at one of three levels of “coupling” with selection. This means that, in one condition, the chance of missing the target was strongly coupled with the selected target, in another, the chance of hitting was less dependent on selection. This allows us to examine how this coupling of selection to execution errors affected the behavioural response to execution error outcomes. A high coupling of execution errors with target selection led to execution errors (on the target that was designed to be difficult to hit) inducing switching behaviour, whereas previous results indicated execution errors induced reselection behaviour. Switching behaviour following an execution error perhaps highlights that gating is not occurring (or occurring to a lesser extent), since it shows the lack of reward is being considered. The shifting of the gating behaviour associated with this coupling implies that this credit assignment process is mediated by a belief about the task structure, as has been recently studied (Moran et al., 2019).

A series of basic reinforcement learning models were used to assess the credibility of different mechanism formulations. The best-fit model parameters were inferred from the data using Bayesian techniques. In each case, a gating hypothesis was broadly supported across these three experiments, as the data supported an “update rate” lower than 1. However, the best performing specific model was different in each case, making it difficult to conclude the exact nature of the true mechanism. This work presents further evidence for the gating mechanism seemingly present in motor credit assignment (McDougle, Boggess, et al., 2016).

Following this work, a more fundamental question about this proposed gating mechanism remained: Is this behaviour a unique product of an interaction between decision making and motor control processes? Perhaps, any multi-stage task where ambiguity about the source of the error is able to be resolved can facilitate the types of behaviour that have been observed in these experiments. To test this, a 2-stage decision making task was developed. In it, the participants must select one of two treasure chests (as they did in the original motor task), but before they gain access to the chest, they must find a key hidden inside one of two cups. The cup guessing stage emulated the motor control component; it was designed to be perceived as being independent of selection (i.e. errors that occur in this stage cannot be attributed to poor selection). The task used identical probabilities as the first experiment in Chapter 6, and ultimately revealed the same patterns of behaviour. This highlights a mechanism broader than gating, the data shows how ambiguity in the structure of the task can lead to incorrect attribution of errors. Specifically, participants generalise any credit they assign to the cup guessing stage between the two chests – presumably because the task is presented in such a way that implies the target selection does not affect the change of finding the key in the cup guessing stage. A model was constructed to formalise this mechanism, and the model fit revealed evidence for parameters which back up the claim that credit was generalised between the two chests in the cup guessing stage.

In general, these findings highlight how a top-down model of the task structure can influence the model-free credit assignment that occurs during learning (Moran et al., 2019). Interactions between model-based and model-free learning, and multi-stage decision tasks are highly topical within cognitive science today (Shahar, Hauser, et al., 2019; Shahar, Moran, et al., 2019), and this work contributes by examining the contributions of sensorimotor control in decision making.

References