THE EFFECT OF REWARD ON MOTOR ADAPTATION AND MOTOR CONTROL

by

Olivier Eugene Georges Codol

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Psychology

The University of
Birmingham
July 2019

Abstract

The prospect for rewarding outcomes has long been known for its impact on human behaviour, and motor control is no exception to this. Recent years were marked by widespread interest in how reward alters motor learning and motor control in humans, and subsequent efforts produced a wealth of descriptive reports underlining which behaviours are shaped by it. More recently, the focus is shifting toward asking which underlying mechanisms drive these alterations and this work adheres to this effort. This thesis is divided into two main parts. First, investigating what underlying mechanisms drive enhancement of motor learning with reward, we see that explicit control is tightly coupled with reward processing in motor adaptation. Extending these findings, we explore which individual characteristics predict sensitivity to reward during motor learning, and observe that working memory, rather than genetic profile, shapes this variability. In the second part, we turn to motor control, and see that enhanced control during reaching is driven by regulation of arm stiffness, in addition to other proposed mechanisms such as feedback control. Finally, in an attempt to manipulate reward-based effects using transcranial magnetic stimulation of the ventromedial prefrontal cortex and supplementary motor area, no alteration of behavioural enhancements was observed.

Acknowledgements

What would even the keenest graduate student be without the helping presence of his supervisor? Fortunately, I have no answer to this question, because I certainly have been blessed with an outstanding supervisor myself. Joseph Galea's constant availability and open-door policy helped me feel comfortable and safe in the workplace as I grew in independence. He left me enough lee-way to experiment in my scientific approach while providing regular feedback to ensure I don't get lost in the woods.

To this day, I can still hardly believe that I was blessed not only with a great supervisor but also with the most caring mentor one can hope for. Peter Holland proved the most joyful debating companion, and our endless discussions about every tiny details of this thesis were possibly one of the best memories I will keep of this period.

Joe and Pete were by far the most pedagogic, efficient and complementary team I could have dreamt for. Joe gifted me with his experience, Pete with his time, and both of them with an endless well of patience and support. I would consider myself lucky to walk away with even a tiny fraction of the scientific insight and quality of character that I witnessed from them.

The remarkable support of my two secondary supervisors should also be mentioned. Jeremy Wyatt's guidance during my first year literally opened me to the wonderful world of optimal control theory, Kalman filtering and virtually anything based on calculus. In that sense, he has had an indelible impact on my work for the years to come. Chris Miall's kindness, passion and scientific guidance throughout also proved greatly invaluable and helped me get a deeper grasp of my work.

I am most indebted to John-Stuart Brittain for his availability and the enthusiasm with which he guided me through challenging mathematical and analytical endeavours. Similarly, my debt extends to Sanjay Manohar, who's help and insight completely reshaped my understanding of chapter 4. I am also greatly thankful for Raphael Schween's friendly discussions on Chapter 2.

I would like to thank all the other Ph.D. students and post-doctoral researchers who made this journey unforgettable: Diar, Ed, Rachel, Wilf, Roya, Seb, Xiuli and Mike; My friends Selim, Cosmin, Mirela and Rahmi, and those who would not let me go despite the distance, Panos, Aditya and Yuto.

Je ne pense pas qu'il y ait de mots pour décrire le sentiment de gratitude que je porte envers mes parents. Ayant eux-mêmes souffert le dénuement pour mener leurs études, ils n'hésitèrent pas à s'expatrier pendant six ans pour s'assurer que mes sœurs et moi-même puissions finir les nôtres dans le confort. À mon père, Roland, et ma mère, Christine, plus que tout autre, cette thèse est la vôtre, car votre passion à entretenir mon acharnement compulsif au débat stérile est sans aucun doute la fondation de mon raisonnement critique – et de mon indubitable érudition concernant l'œuvre de Ronsard. En ce sens, mes sœurs Fanny et Florence se sont avérées de féroces et vaillants compagnons de séance. Je remercie ma bonne étoile tous les jours d'avoir une si belle famille.

Enfin, je dois mon éternelle reconnaissance à mes plus chers amis Guillaume, Loïc et Christophe pour leur calleuse camaraderie qui n'a que faire du temps et de la distance.

Contents

List of Figures

List of Tables

Acronyms

AMT active motor threshold.

ANOVA Analysis of Variance.

BF binary feedback.

BIC Bayesian information criterion.

CCW counter-clock wise.

CI confidence interval.

COMT Catecholamine-O-Methyl-Transferase.

DARPP32 Dopamine- and cAMP-Regulated neuronal Phosphoprotein.

DRD2 Dopamine Receptor D2.

FRT fast reaction time.

M1 the primary motor cortex.

MAE mean absolute error.

MRI magnetic resonance imaging.

MT movement time.

NHST null-hypothesis significance testing.

PD Parkinson's disease.

PFC prefrontal cortex.

RT reaction time.

RWM mental rotation working memory.

SMA the supplementary motor area.

SNP single nucleotide polymorphisms.

SRT slow reaction time.

SWM spatial working memory.

TMS transcranial magnetic stimulation.

VF visual feedback.

vmPFC ventro-medial prefrontal cortex.

VWM verbal working memory.

WM working memory.

WMC working memory capacity.

Chapter 1

INTRODUCTION

1.1 Motivation of this thesis

Movement is the main way animals interact with their immediate environment. However, controlling the motor system proves an outstandingly difficult problem. An infinite variety of situations with different and often complex constraints can arise in the environment, leading to ever-changing requirements for optimal behaviour. This makes learning from new situations a critical process dictating how efficient a system or individual is at controlling movement. It is therefore only natural that training is a central part of any motor skill, and mankind has always thrived to find methods to improve learning speed or strengthen the memory of acquired motor skills over longer periods of time.

To this end, reward proves a timeless centre of interest. For instance, it is well-known that reward-based episodic memories (*e.g.* using images of appealing food or landscapes) are much more strongly remembered (Hamann et al., 1999) even after many years, and that reward serves as a central source of information for defining new behavioural policies when faced with decision-making problems (P. Dayan & Daw, 2008). Much research has focused on shaping a better understanding of the exact role of reward in neuroscience, and the field of motor control is no exception to this. The recent years has seen a rising interest on this matter in motor adaptation, a sub-field of motor learning research (Haith & Krakauer, 2013; V. S. Huang et al., 2011; Izawa & Shadmehr, 2011). One reason motivating this interest is the potential for enhancing rehabilitation procedures for clinical populations that underwent a loss in motor abilities (Chen, Holland & Galea, 2018; Goodman et al., 2014; Quattrocchi et al., 2017). This present work focuses mainly on further extending our understanding of the nature and effect of reward on motor learning and motor control.

1.2 Current views on how the brain controls the body

Learning is by essence a study of change, and understanding a learning process naturally requires understanding what is being altered. For motor learning, the system being changed is the motor control system, which is also occasionally referred to as the "controller" from the biomechanics literature. Therefore, in this section, we will discuss the current consensus on how the brain implements motor control at a systems level.

1.2.1 Inverse model and feedforward control

To perform any action, the body part(s) involved—also called the effector(s)—must receive "motor commands" descending from the brain in the form of patterns of nerve impulses activating muscle contractions and leading to the action itself. These motor commands can be thought of as a series of muscle activation signals. What set of motor commands lead to a specific action is a computationally tricky question for at least two reasons. First, there is no single and unique solution to this problem, that is, it is a redundant system (Hirashima & Nozaki, 2012). For instance, reaching movements to an object generally occurs in 3 dimensions, but the arm contains seven degrees of freedom, therefore allowing many different reaching movements to be possible (V. Martin et al., 2009). Second, it requires having an inverse model of the effector (figure 1.1), sometimes called inverse model of the plant (Bhushan & Shadmehr, 1999). An inverse model can be conceptualised as a function that gives the motor command u given the desired state x^*. It is called an "inverse" model because it is the inverse of the function that gives an action x^* when given the motor command u, which is merely a model of the effector itself. In simpler words, an inverse model is the solution, or a solution, to the problem of finding the motor commands u that will make the effector perform the action leading to the state x^*.

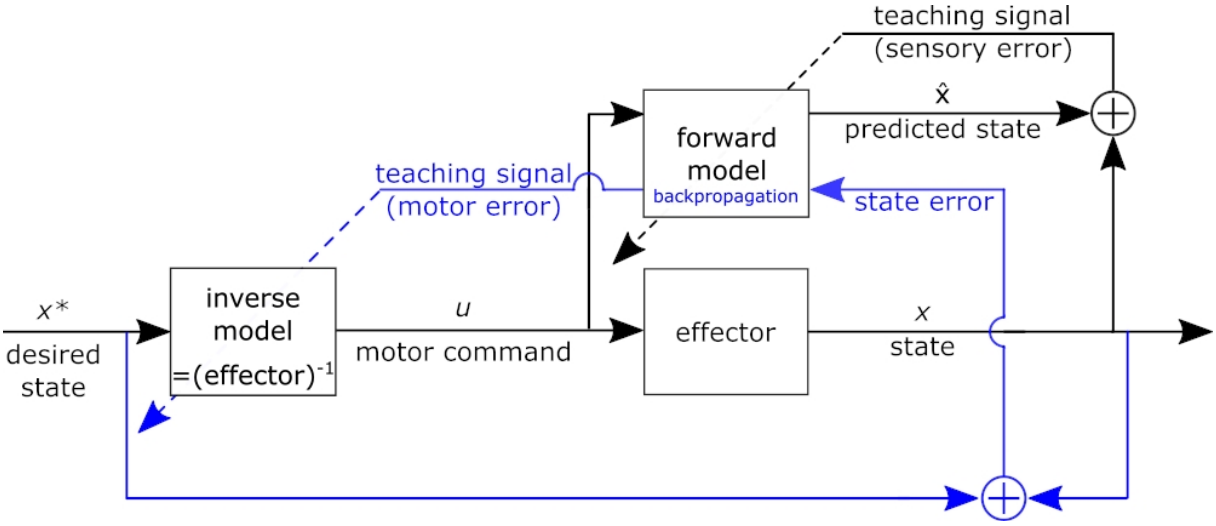


Figure 1.1: Schematic of a feedforward controller using an inverse model and a forward model. The blue component indicates inverse model learning, which is enabled by the existence of a forward model (M. I. Jordan & Rumelhart, 1992; Wolpert & Miall, 1996).

Inverse models are powerful controllers because they can reliably provide a suitable solution for any biomechanically possible desired action (Wolpert & Kawato, 1998; Wolpert et al., 1998). However, there are several issues with this type of controller. First, it is feedforward, meaning that it produces motor commands independently of the actual movement and its potential deviation from the desired trajectory. This is incompatible with a body of experimental data showing that corrective control does occur (Bhushan & Shadmehr, 1999; Carroll et al., 2019; Kasuga et al., 2015; Scheidt et al., 2005; Tseng et al., 2007). Second, learning an inverse model proves challenging, because no reference of what the correct output should be exists to serve as a teaching signal; if there was any, there would be no need for an inverse model in the first place (Wolpert & Kawato, 1998; Wolpert et al., 1998). Though plausible solutions have been found to overcome this issue and allow for learning at the inverse model level (Gomi & Kawato, 1993; Wolpert & Miall, 1996), it remains that learning at the inverse model stage is too slow to account for observations in human motor learning (Bhushan & Shadmehr, 1999). Rather, the consensus on this issue is that the brain controls movement through a combination of inverse and forward models (Honda et al., 2018). Together, inverse and forward models form a broader category of so-called "internal models".

1.2.2 Forward models and feedback control

What is a forward model, and why it is necessary for control

A forward model predicts the expected resulting state $\hat{x}_{t+1}$ following the motor command u being applied to the effector (figure 1.1), granted it knows the initial state x_t before u is applied. This additional step conveniently augments an inverse model by overcoming some of the issues mentioned in the previous section (Bhushan, 1998).

First, forward models enable feedback control (Wolpert & Kawato, 1998). They can easily track deviations from the desired trajectory, by comparing sensory feedback from the real world with forward model predictions. A standalone inverse model is unable to determine deviations from trajectory, because sensory information signalling deviations from a planned movement are tangled with sensory feedback from the planned movement itself. While an inverse model can predict a required action to reach a sensory state, a forward model can predict what sensory state is expected by the planned movement, using forward computation ("what feedback will result from these motor commands coming down to the effector?"). Therefore, by comparing predicted and actual feedback, forward models can isolate the sensory consequences of external perturbations at any time during the reach, while an inverse model alone could only compare a desired final state with the actual final state and initiate a follow-up movement afterwards (Bhushan & Shadmehr, 1999; Shadmehr & Krakauer, 2008; Wolpert et al., 1998).

A critical issue of feedback control is that it involves using a sensory feedback signal, which usually includes delays related to carrying information from the sensors in the effector (*e.g.* tactile nerve endings) to the central nervous system. These delays expose the system to instability issues. However, forward models can provide part of the solution, by estimating the current state of the effectors based on the last available sensory information and motor commands (Miall et al., 2007). This estimate can then inform the inverse model of the expected current state of the motor system, enabling it to perform feedforward computation in an optimal fashion by overcoming sensory delays. It is important to distinguish this contribution from detecting deviations from desired states, in that even in the absence of any external disturbance, the sensory feedback delays must be accounted for to promote optimal control (Miall et al., 2007).

Finally, deviation from a predicted state can be used as a sensory prediction error signal (Haruno et al., 2001; Mazzoni & Krakauer, 2006; Tseng et al., 2007), which can easily train a forward model using a supervised update rule (Thoroughman & Shadmehr, 2000). This means that learning is very easy to implement and therefore much faster in a forward model than in an inverse model, making the controller much quicker to adapt to novel environmental constraints (Bhushan & Shadmehr, 1999). This property enables powerful adaptive behaviour such as those observed during visuomotor adaptation tasks, which we detail later.

Therefore, the advantages of a forward model include:

- Enabling feedback control by isolating expected consequences of motor commands from consequences due to external perturbations at any time.

- Cancelling out sensory feedback delays in the sensorimotor system by predicting the current state of the effector.

- Faster learning upon exposition to unexpected systematic perturbations using sensory prediction errors.

The latter point especially will have a pivotal role in this work and will be discussed in more detail in section 1.3.

Evidence of the existence of a forward model in humans: the cerebellum

While the exact localisation in the brain of inverse models has long been an elusive question, recent empirical and computational evidence point toward either a cerebellar localisation (Alvarez-Icaza & Boahen, 2012) or a shared motor cortex-cerebellum localisation (*e.g.* Honda et al., 2018). In contrast, the cerebellar localisation of forward modelling is less debated due to an extensive series of evidence. For instance, clinical populations characterised by cerebellar dysfunction (cerebellar patients) express deficits in online control of movements (Holmes, 1939; Tseng et al., 2007; Vilis & Hore, 1980). Furthermore, the capacity to account for delays in sensory feedback is near non-existent in cerebellar patients, as well as in healthy participants that underwent focal and transient disruption of cerebellar activity following transcranial magnetic stimulation (TMS) procedure (Miall et al., 2007). Causal evidence is also found in monkeys with focal disruption of the cerebellum dentate and interpositus nuclei via cooling techniques (Vilis & Hore, 1980). This manipulation led to cerebellar tremor during reaching movements, showing that motor control is more prone to instability in the absence of predictive information from the cerebellum. Finally, cerebellar patients express strong learning deficits when adapting to new environmental disturbances (Tseng et al., 2007). For instance, adapting to a visual shift using prism glasses during a ball-throwing task proves nearly impossible for cerebellar patients, while healthy patients express relatively quick adaptation (T. A. Martin et al., 1996).

Feedback control

Different forms of feedback loops contribute to motor performance. The fastest feedback loop is the spinal reflex, or stretch reflex, which uses proprioceptive information to trigger a corrective response originating from the spinal cord to bring the effector back to its original position after a displacement. It can usually be measured within 25-50ms after perturbation onset (Weiler et al., 2019). The transcortical reflex is slightly slower, acting around 50-100ms after a perturbation because sensory information passes through the primary motor cortex (Pruszynski et al., 2011). Finally, visuomotor feedback makes use of visual information to adjust a movement based on visual information. Because it requires higher levels of integration, it is slower, usually showing delays of at least 170ms (Carroll et al., 2019). Each of these systems differ not only in their delays, but also in their level of integration and the sensory sources being integrated, resulting in a system that displays some level of redundancy but also some level of specialisation (Carroll et al., 2019; Omrani et al., 2016; Weiler et al., 2019).

1.2.3 Motor Planning

While motor planning is an entire subfield in its own right, for this work we are mainly concerned about the problem of action selection. The planning stage of motor control usually refers to the stage directly preceding movement itself during which the movement is prepared but no movement is performed (Churchland et al., 2006). This includes the process of selecting or “choosing” an action to achieve a desired goal amongst a set of possible actions (Gurney et al., 2001). The more complex a “choice” is, the more computation will be required and the longer the movement reaction time will be (Haith et al., 2015, 2016).

1.3 Motor learning processes

A critical issue in motor control is that neither the body nor the environment is stable over time: riding a new bicycle or using a baseball bat with a different weight distribution for instance will unpredictably alter environmental consequences of the same action, while one's body changes throughout one's lifespan, constantly invalidating the accuracy of current internal (*i.e.* inverse and forward) models. Therefore, accounting for those changes on the basis of motor errors is a critical factor defining the effectiveness of the controller itself.

1.3.1 Learning *de novo* versus motor adaptation

Two types of motor learning can be defined. First, when performing a new task, if adequate inverse and forward models are not already acquired, one must usually acquire those models "from scratch". This is often referred to as learning *de novo* (Kasuga et al., 2015; Telgen et al., 2014). This is to be distinguished from another form of learning, where a relatively close set of inverse and forward models already exist for the motor task considered, while not being accurate enough to lead to satisfying performance. This second form of learning is called motor adaptation, because it merely requires adjusting the pre-existing internal models without implementing any structural change (T. A. Martin et al., 1996; Shadmehr & Mussa-Ivaldi, 1994; Tseng et al., 2007).

Within motor adaptation, three contributing mechanisms have drawn widespread interest. Forward model recalibration driven by performance error has historically drawn the most research effort. In addition to forward model recalibration, which can be rather slow (T. A. Martin et al., 1996; Shadmehr & Mussa-Ivaldi, 1994), humans can also express a much faster, more volitional form of error-reduction approach called explicit control (Morehead et al., 2015; Taylor et al., 2014; Taylor & Ivry, 2011, 2014). In this section we will specifically examine these two mechanisms and empirical evidence toward each of them. Finally, reinforcement-based (reward-based) learning has been more recently put forward as another candidate mechanism (Haith & Krakauer, 2013; V. S. Huang et al., 2011; Izawa & Shadmehr, 2011; Taylor & Ivry, 2014). This mechanism is the pivotal point around which this work will evolve and will be discussed in more detail later.

1.3.2 Motor adaptation

Sensory prediction error as a teaching signal

As mentioned earlier, inverse models are less susceptible to fast and efficient learning when compared to forward models (Gomi & Kawato, 1993; Wolpert & Kawato, 1998). Therefore, it has been suggested that when adapting to new environmental constraints, motor adaptation occurs mainly at the forward model stage. This proposition has been supported by comparisons of human reaching data with simulation outputs from various possible controller architectures (Bhushan, 1998) and behavioural data manipulating feedforward and feedback control availability during an adaptation task (Tseng et al., 2007). This suggests that sensory prediction error is the key signal driving motor adaptation, because it is the signal driving learning in forward models (Mazzoni & Krakauer, 2006; Tseng et al., 2007). Thus, although it is likely that the inverse model plays a role by adapting at a slower pace (Kawato & Gomi, 1992; Honda et al., 2018), it is common and an accurate enough approximation to consider that internal model recalibration is driven by sensory prediction errors alone. This is also commonly referred to as cerebellar adaptation, due to the well-established cerebellar localisation of forward models (Miall et al., 2007; Shadmehr & Krakauer, 2008; Tseng et al., 2007; Wolpert et al., 1998).

In laboratory conditions, motor adaptation is often studied using a paradigm called visuomotor adaptation (Mazzoni & Krakauer, 2006; Morehead et al., 2017; Tseng et al., 2007; figures 1.2A and 1.3A). In this now seminal paradigm, participants perform reaching movements from a starting position to a target, while their hand is hidden away from their view. Rather, they are informed of their hand position through a cursor on a screen (online visual feedback), displayed alongside the target and starting position. At some point, a visuomotor displacement—also sometimes called visuomotor rotation—is introduced, effectively rotating the reaching direction of the cursor by a fixed angle (figure 1.2B). Consequently, reaching errors arise that were not predicted by the controller, leading to a sensory prediction error and enforcing adaptation within the context of that task (figure 1.3B). The simplicity and flexibility of this task made it a very powerful approach to test the characteristics and limitations of motor adaptation in humans.

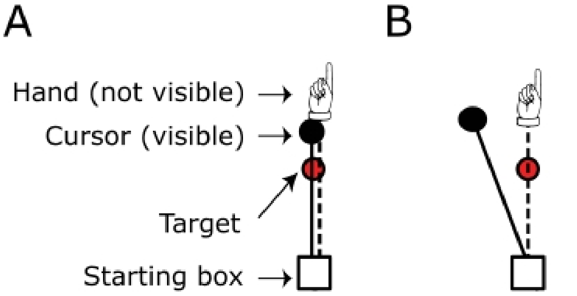


Figure 1.2: Illustration of a visuomotor rotation. A. In the baseline condition, participants reach to a target with a cursor that follows their hand position. B. when a visuomotor displacement is introduced, the cursor follows the hand position with an angular deviation from the hand trajectory.

Fixed learning rates during cerebellar adaptation

How are sensory prediction errors driving adaptation of forward models? A very simple learning rule for forward models is the gradient descent learning rule. This is a form of supervised learning that uses error from a reference signal (here the sensory prediction error) to push the model in the direction opposite to the error (Izawa & Shadmehr, 2011; Shadmehr & Krakauer, 2008; Thoroughman & Shadmehr, 2000), giving:

$$\hat{x}_{t+1} = \hat{x}_t + \gamma \cdot (y - \hat{y}) \tag{1.1}$$

where $\hat{x}_t$ is the model's state prediction for trial t, and y and $\hat{y}$ are the observed and predicted sensory feedback. Possibly the most important feature of this learning rule is the learning rate parameter γ. If the learning rate is too small, learning will be slow and the system too rigid over time; if it is too high, the system will overlearn and large jumps in performance will occur, leading to instability. Experimentally observed adaptation profiles in visuomotor adaptation show that learning rates are actually quite steady across a large range of error magnitudes (Morehead et al., 2017; figure 1.3D).

Aftereffects

Once full adaptation has been reached, removing the rotation to reintroduce veridical feedback shows the existence of "aftereffects" (Kitago et al., 2013; Morehead et al., 2017; Tseng et al., 2007): despite the return of normal feedback, participants now express errors in the direction opposite to the rotation, by still reaching to the direction that previously successfully accounted for the visuomotor rotation (figure 1.3B,C). This aftereffect-related error then drives recalibration back toward baseline (Kitago et al., 2013), and so the aftereffect quickly fades away over trials. Aftereffects are a very practical way of measuring post-adaptation levels of forward model recalibration for a given participant because it is a completely implicit process, and is therefore not corrupted by explicit control, as will be discussed later (Bond & Taylor, 2015; Morehead et al., 2017; Taylor & Ivry, 2011).

Limits of cerebellar adaptation: forgetting

While cerebellar adaptation can account for a large variety of perturbation magnitudes, learning usually saturates after 10° to 20° (Bond & Taylor, 2015; Haith et al., 2015; Huberdeau et al., 2015; Leow et al., 2017; Morehead et al., 2017; Telgen et al., 2014; figure 1.3B). Why forward model recalibration is limited in such a way is still an open question, but a possible explanation is that a dose of forgetting occurs concurrently to learning (Cheng & Sabes, 2006; Morehead et al., 2017; Thoroughman & Shadmehr, 2000).

Once learning and forgetting reach a null net contribution, one would cancel out the other, leading to saturating levels of adaptation. Though this has not yet been proved to occur, the main strength of that hypothesis is that it explains the invariance of saturation levels across a wide range of sensory prediction errors (Bond & Taylor, 2015; Morehead et al., 2017).

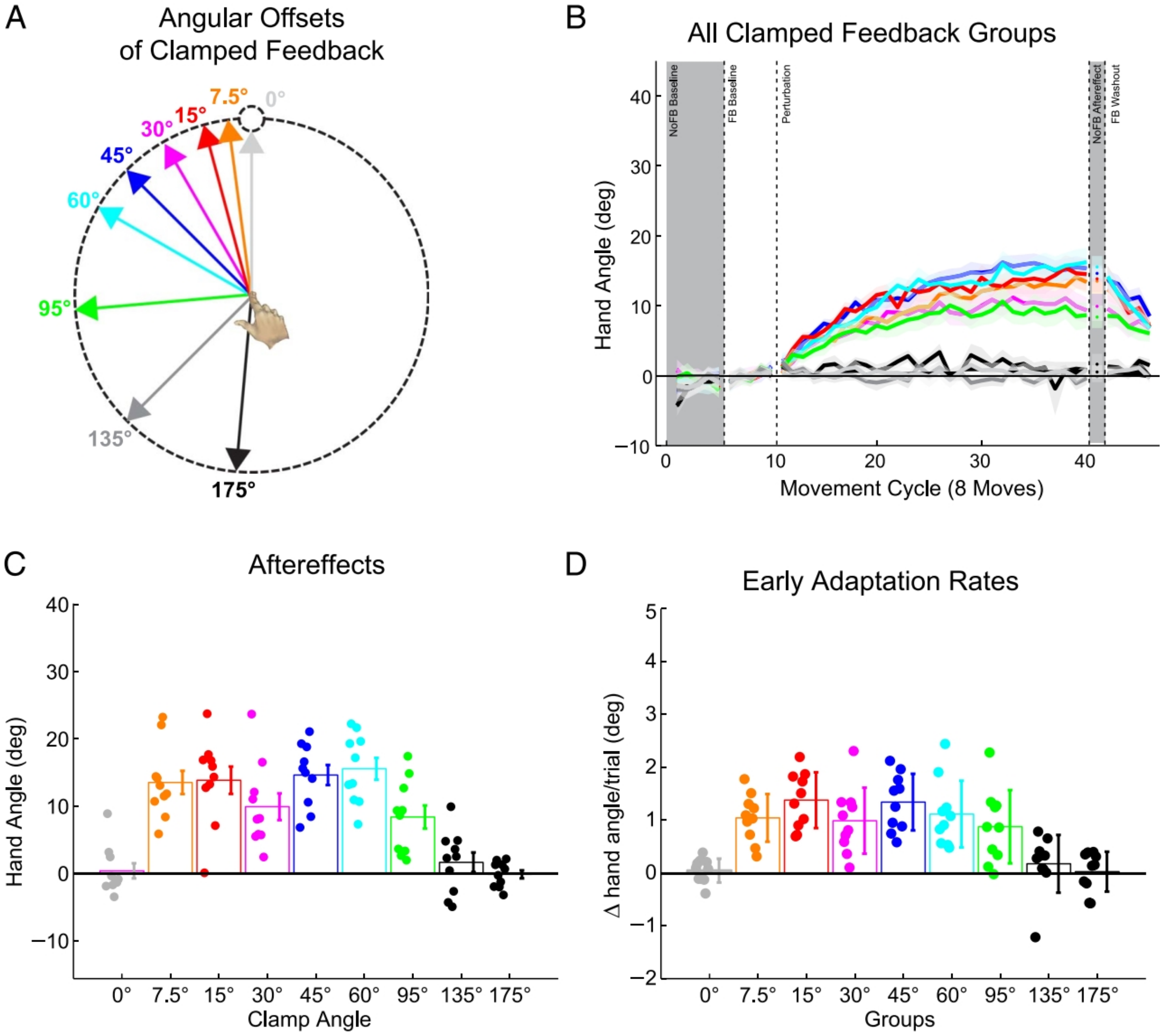


Figure 1.3: Characteristics of cerebellar adaptation. Figure adapted from Morehead et al. (2017). A. Participants were exposed to visuomotor rotations of different magnitudes. B. Reaching performance. Vertical dashed lines indicate block delimitations. After an initial baseline block with no visual feedback and another with veridical feedback, participants were exposed to a visuomotor rotation and reach angles adapted accordingly. Aftereffects were measured using a short no-feedback block, and veridical feedback was then reintroduced in the last block. C. Aftereffects were similar across all displacement magnitudes below 95°. D. Learning rates were similar across all displacement magnitudes of less than or equal to 95°.

What evidence is there of forgetting? To answer this question, an important distinction must be made between two different forgetting processes. First, extinction is a well-documented phenomenon in the saccade adaptation literature (De Zeeuw & Ten Brinke, 2015; Jirenhed et al., 2007), whereby the memory of a previously learnt perturbation rapidly decays when one is exposed to the associated context without any teaching signal to maintain the acquired memory. In the context of visuomotor adaptation, one can learn the visuomotor displacement at first, but once visual feedback (*i.e.* the cursor—which serves as a teaching signal) is removed, the memory should therefore decay because contextual information remains the same (*i.e.* apparatus, target positions, etc.), exposing the memory to extinction. Indeed, this has been observed empirically numerous times (Galea et al., 2011, 2015; Kitago et al., 2013).

Second, even in the absence of contextual input, time alone can explain a decay of synaptic memory in the cerebellum (S. Kim et al., 2015; Kitago et al., 2013; Lago-Rodriguez & Miall, 2016). Electrophysiological evidence comes from cerebellar Purkinje cell recordings in awake monkeys during smooth pursuit tasks. The authors found that a change in activity following learning in one trial was forgotten after 6s (Yang & Lisberger, 2014). Behaviourally, manipulating inter-trial time intervals in a visuomotor adaptation task also shows that the higher the interval between trials, the less learning takes place (S. Kim et al., 2015).

Cerebellar adaptation: summary

Overall, a large body of work on cerebellar adaptation shows that it is an outstandingly stereotyped phenomenon, even across individuals, a characteristic that is somewhat surprising considering that humans have a strong tendency to express a wide range of idiosyncratic behaviour in most tasks. The main reason behind this is likely that idiosyncrasies are a direct consequence of the great variety of strategic approaches (explicit control) employed depending on people, while cerebellar adaptation is an implicit, strategy-free process (Bond & Taylor, 2015; Haith & Krakauer, 2013; Morehead et al., 2017; Taylor & Ivry, 2011, 2014). In the next section, we will discuss the role of explicit control in motor adaptation, and how it interacts with forward model recalibration.

1.3.3 Explicit control

Forward model recalibration occurs independently of explicit control

One of the early questions regarding explicit control is to what degree it influences forward model recalibration. In a now seminal study, (Mazzoni & Krakauer, 2006) asked participants experiencing a visuomotor rotation to reach not for the target, but off from the target in order to counter for the displacement (figure 1.4A). Effectively, this means that participants' explicit control was clamped to a contribution equal to the displacement in magnitude and opposite to it in direction, immediately compensating for it and preventing any task error from arising. Nevertheless, participants slowly drifted away from the aiming-off direction, despite them reporting trying to maintain their reaching direction constant according to instruction (figure 1.4B). This drift was in accordance with the discrepancy between their actual reaching movement and the visually displayed outcome movement, clearly suggesting that forward model recalibration was taking place regardless of the absence of any task error. This study was taken as proof that forward model recalibration takes place regardless of the presence or absence of explicit control and that it cannot be prevented by participants, *i.e.* it is independent, automatic, and implicit. Later, further reports confirmed this hypothesis (Bond & Taylor, 2015; Morehead et al., 2017), though it has been shown that visual landmarks manipulating uncertainty of explicit strategies can alter cerebellar adaptation (Taylor & Ivry, 2011).

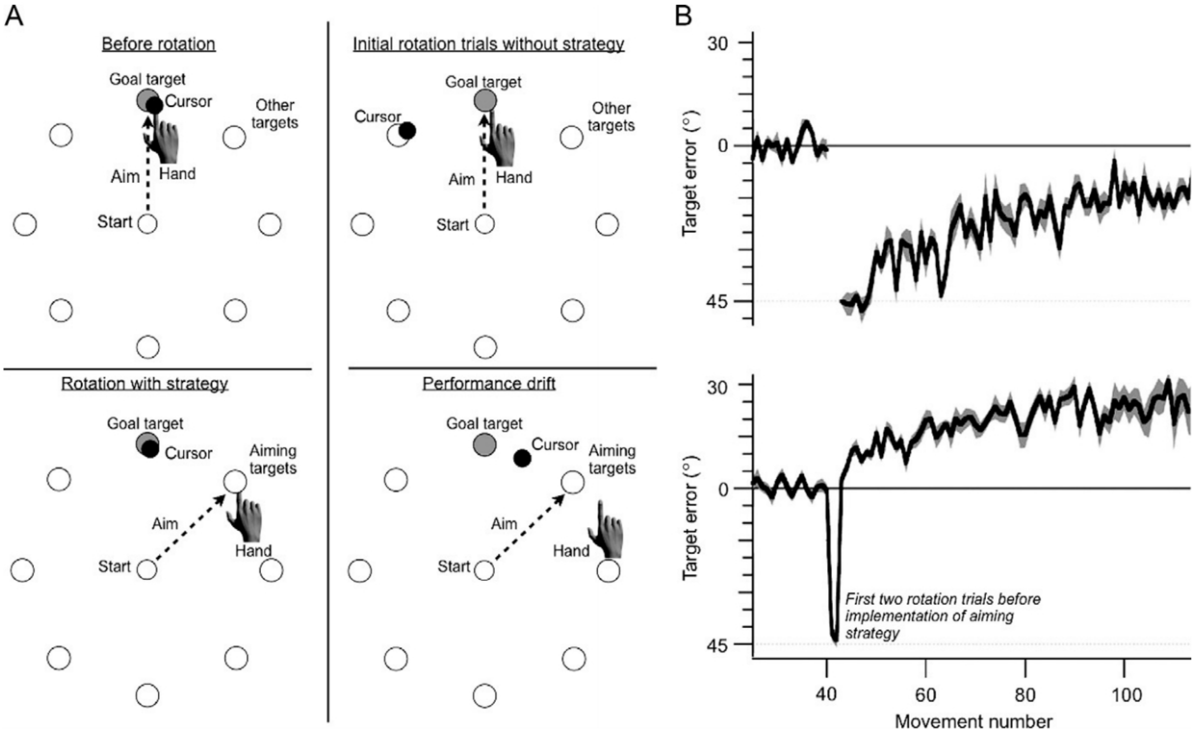


Figure 1.4: Implicit adaptation occurs independently from explicit control. Figure adapted from Taylor & Ivry (2014), original experiment from Mazzoni & Krakauer (2006). A. Top left: baseline trials with veridical feedback; participants reached to a set of 8 targets. Top right: a 45° visuomotor rotation is introduced. Bottom left: Participants are told to aim off from target by 45° to counter the rotation. Bottom right: Even though task error was null on average, participants still drifted away from target, showing implicit adaptation. B. Reaching performance across trials. The top panel shows a normal adaptation profile, where participants are not told to use any strategy to counter for the displacement. The bottom panel shows performance when participants are told to use a strategy. A drift can still be observed over trials.

Fast explicit control and slow cerebellar adaptation

The study from Mazzoni & Krakauer (2006) presents one shortcoming: it effectively clamps explicit control and thus prevents us from observing the interplay between both components over time. To that end, a long-standing issue has been the concomitant expression of explicit control with forward model recalibration, obscuring the individual contribution of each (Taylor & Ivry, 2014). Fortunately, the past ten years have seen great progress in this regard, as several different paradigms were proposed to quantify each component concurrently. A first paradigm involved asking participants to indicate their planned reaching direction on each trial (Bond & Taylor, 2015; Taylor & Ivry, 2011; Werner et al., 2015). The discrepancy between planned reach direction and actual reach direction was considered as the contribution of forward model recalibration. Another approach takes advantages of the cognitively demanding nature of explicit control (Anguera et al., 2012, 2010; Benson et al., 2011; Huberdeau et al., 2015), and enforces early initiation of reaching movements before any explicit contribution can take place (Fernandez-Ruiz et al., 2011; Haith et al., 2015; Leow et al., 2017). Interleaving early initiation and normal initiation trials allows sampling the full contribution of both explicit control and forward model recalibration in one trial and the sole contribution of forward model in a subsequent trial.

Dissociating explicit control contribution from cerebellar adaptation shows that early performance is mainly driven by explicit control (Bond & Taylor, 2015; Huberdeau et al., 2015; Morehead et al., 2017; Taylor & Ivry, 2011). This is because cerebellar adaptation occurs more slowly, and therefore any early attempt to reduce systematic errors must rely on explicit control. However, as more of the perturbation becomes accounted for by forward model recalibration, explicit control becomes redundant, and even counter-productive (Mazzoni & Krakauer, 2006). Therefore, explicit control contribution decreases proportionally to the length of exposition to the disturbance (Bond & Taylor, 2015; Taylor & Ivry, 2011). In other words, the explicit control contribution accounts for the proportion of the displacement that cerebellar adaptation does not account for (figure 1.5).

What triggers explicit control contribution?

The flexibility that explicit control exhibits suggests that its involvement is not only varying over time, but that it is also sometimes non-existent. Therefore, what is the mechanism behind its recruitment, or its suppression? Evidence suggests that during visuomotor adaptation, the temporal schedule of displacement introduction is central to promote or prevent recruitment of explicit control. For instance, encountering large errors

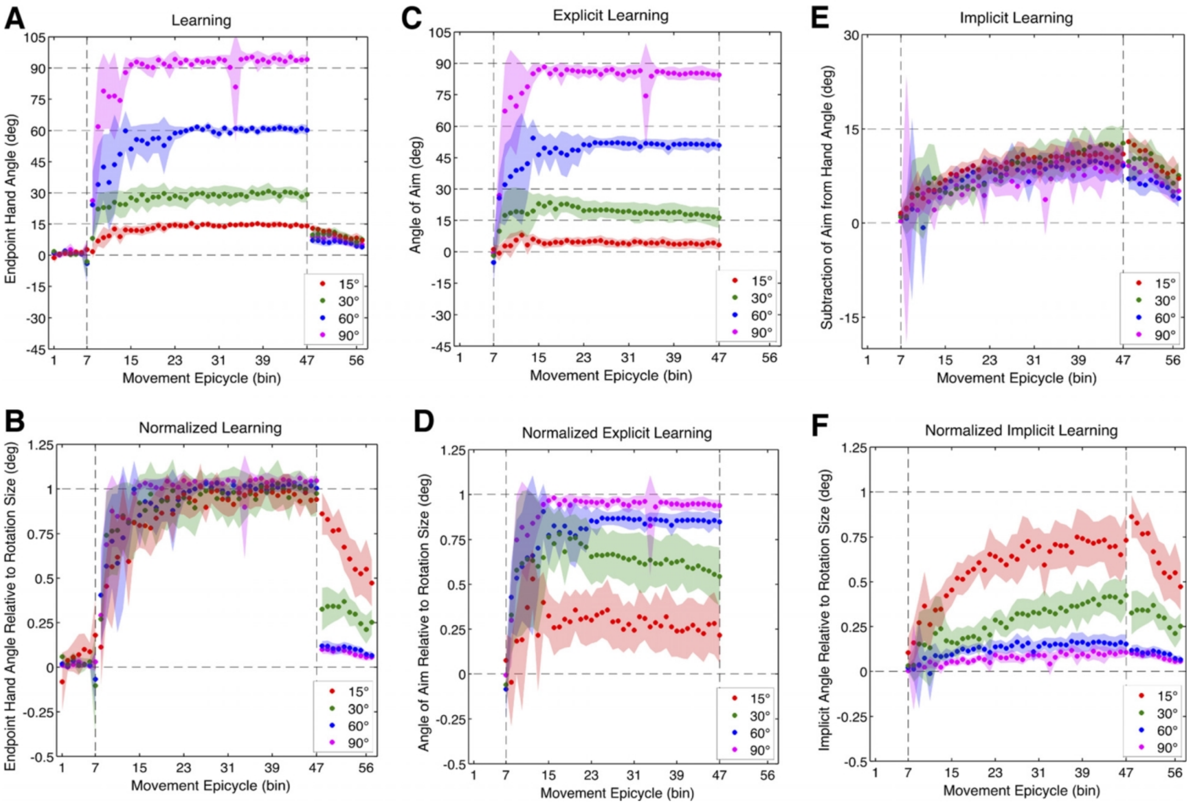


Figure 1.5: Contribution of explicit control to visuomotor adaptation. Figure adapted from Bond & Taylor (2015) A. Reach angles for fifteen, thirty, sixty, and ninety degrees displacements. B. Normalized learning: end-point hand angle divided by the size of the rotation for each group. C. Explicit learning: angle of aiming location (verbally reported landmark). D. Normalized explicit learning: average angle of aiming location divided by the size of the rotation for each group. E. Implicit learning: subtraction of aiming direction from end-point hand angle. F. Normalized implicit learning: subtraction of aiming direction from end-point hand angle divided by the size of the rotation for each group. Vertical dashed lines denote when the rotation was introduced and removed. Movement epicycles represent the average of an 8-trial bin, and shaded areas represents the 95% CI of the mean.

during a reaching task tends to provoke explicit control (Leow et al., 2017; Malfait, 2004; Werner et al., 2015), because credit is given to environmental factors regarding the underlying cause of this error. Conversely, introducing a displacement with a gradual schedule, so as to prevent exposition to errors beyond a given magnitude, prevents awareness of the manipulation and therefore the involvement of explicit control (Christou et al., 2016; Leow et al., 2016). However, even in the absence of large errors, adaptation leads to awareness as well once it reaches a certain threshold (Bond & Taylor, 2015; Werner et al., 2015), due to the inherent limitation of implicit cerebellar adaptation. Therefore, depending on participants, one should expect a form of explicit control to take place at least after 15° to 20° of adaptation in a visuomotor rotation paradigm employing a gradual displacement.

1.4 Reinforcement learning and motor adaptation

In addition to cerebellar adaptation and explicit control, reinforcement has been proposed to contribute to motor adaptation. Reinforcement is a widely established mechanism in the field of decision-making (Daw et al., 2005, 2006, 2011; P. Dayan & Daw, 2008), where it guides learning of policies based on rewarding feedback (Sutton & Barto, 1998). In short, actions leading to rewarding outcomes see their value increased, while actions leading to punishing or non-rewarding outcomes see their value decreased. The likelihood of selecting an action in the future is then related to the learned value of the set of all available actions. Models of decision-making usually vary in two ways: which rule is employed to update values and which rule is used to select actions based on these values (Sutton & Barto, 1998). These rules are sometimes referred to in the reinforcement literature as the "update rule" and the "policy", respectively.

In this section, we will first discuss some basic concepts from the reinforcement literature, in order to subsequently consider reinforcement-related advances in the field of motor adaptation with a more comprehensive perspective. Finally, we will examine the potential applications of reinforcement in medical procedures such as rehabilitation.

1.4.1 Model-based learning and model-free learning

Two classes of algorithms can be distinguished in the reinforcement literature (Daw et al., 2005; Sutton & Barto, 1998). First, model-free algorithms learn the value of an action by adjusting it as it is used. This means that in order to determine its value, an action must first be expressed, but also that no structural understanding of the task performed is necessary, because the value of all the other, non-selected actions are not updated. This can be seen both as an advantage and a drawback, because it is a very straightforward system to implement but it generalises poorly across the action set, which makes it a slow learning approach. The term "structural understanding of the task" is usually referred to as a "model" of the task (Manley et al., 2014), leading to the name "model-free" for this family of algorithms. Such "structure" or "model" con be conceptualised as the set of relations between all the possible states and actions that the task encompasses, *e.g.* what actions lead to which states, and which states allow to reach another state. On the other hand, model-based algorithms explicitly utilize a model of the task structure in order to update not only the value of the action expressed, but also the value of actions related to it, even though they have not been selected.

To illustrate the importance of this difference, let us imagine we are assessing the bias of a rigged coin by performing a series of throws. If a throw results in a tail outcome, only the probability estimate of the tail outcome increases, while the probability outcome of the head does not. Using a model-based algorithm, the probability estimate of the head should also be decreased, because an accurate model of the task would inform that the tail and head outcomes are mutually exclusive and complementary (*i.e.* if one occurs, the other one doesn't, and one of them must always occur). This is an important distinction: even though the head outcome did not occur, the model-based algorithm is able to update the estimate of that outcome (Daw et al., 2005, 2011; Sutton & Barto, 1998).

Though it looks like model-based approaches are strictly superior, they have also great limitations. They require a full understanding of the task structure, which is often not directly available and must be constructed first (Manley et al., 2014). This process is referred to as structural learning in the motor control literature, though other fields may use different names. Further, even when an accurate model is acquired, working out the relationship between all actions and outcomes based on acquired experience is a computationally intensive process (Daw et al., 2005; Huys et al., 2012, 2015; Otto, Gershman et al., 2013; Otto et al., 2015), which scales very poorly with task complexity: as complexity of a task structure increases, computational requirement of model-based value updates increase even more. This is more often the case than not in motor control, as the environment is generally non-linear, sometimes unstable, time-varying, and can present a large amount of singularities.

This dichotomy between model-based and model-free reinforcement is now well established in decision making. Many studies have now shown that model-based decision-making is a conscious process that relies on working memory to work out task structure and update values (Daw et al., 2005, 2006, 2011; Otto, Gershman et al., 2013; Otto et al., 2015), while model-free reinforcement is a much more implicit learning process (Daw et al., 2005; Dolan & Dayan, 2013; Huys et al., 2012, 2015; Otto, Gershman et al., 2013; Otto et al., 2015). In motor adaptation, however, the idea that reward shapes learning has been introduced much more recently.

1.4.2 Reinforcement in motor adaptation

One of the first papers proposing the existence of a reinforcement component to motor adaptation is a study from V. S. Huang et al. (2011). In this study, participants adapted (marked as Adp^{+} in figure 1.6) to series of visuomotor perturbations with varying target positions and displacement magnitudes (from 0 to 40°), but with a same hand solution (figure 1.6A). In simpler words, the reaching direction required to account for every single visuomotor displacement was always the same, and was therefore repeated (marked as Rep^{+} in figure 1.6). They then assessed participants' memory of the displacement by re-exposing them to each displacement. Results showed a faster memory recollection in this group compared to a control group that adapted to the same targets but did not have the same single solution across all targets ($Adp^{+}Rep^{-}$ group in figure 1.6B). The authors argued that the enhanced memory suggested the existence of a reinforcement

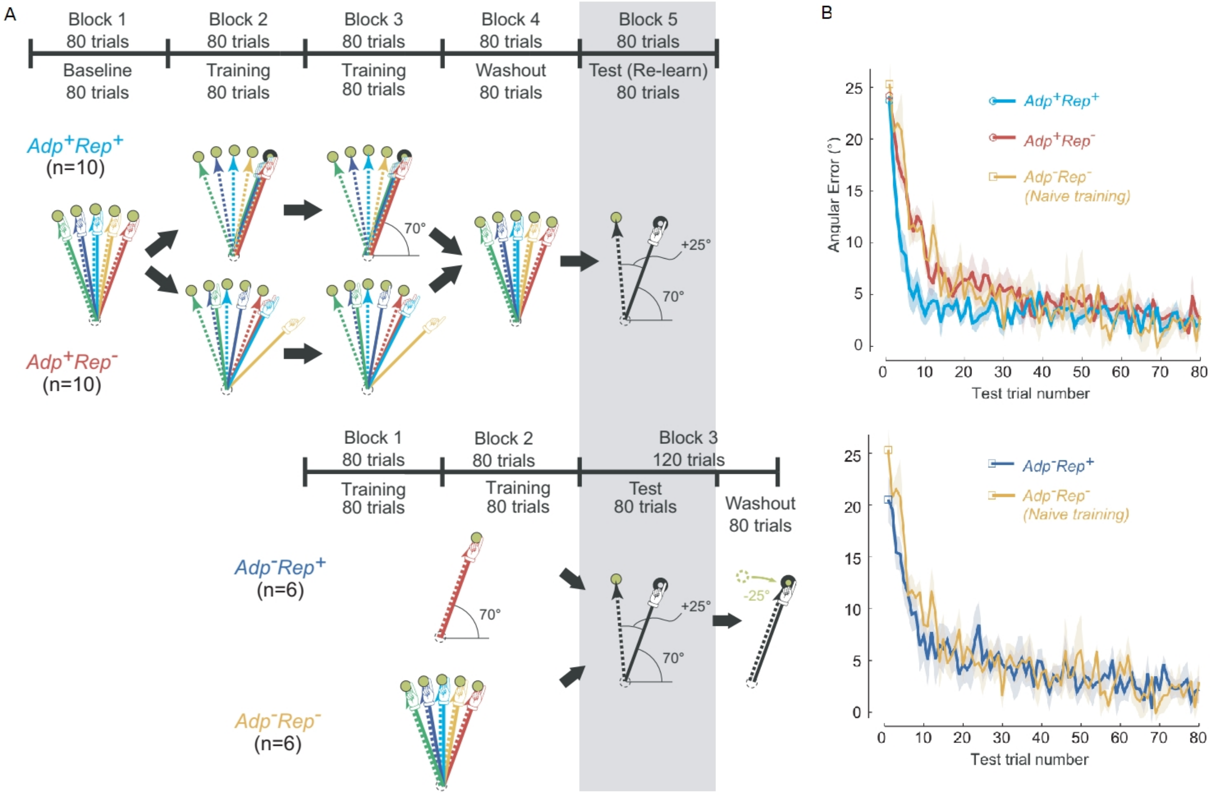


Figure 1.6: Repetition of a rewarded reaching direction led to faster relearning during adaptation. Figure adapted from V. S. Huang et al. (2011). A. Experimental paradigm. After reaching with veridical feedback at a series of 5 targets positioned every 10° over a 40° span, different displacement patterns were introduced for each of 4 groups. In the $Adp^{+}Rep^{+}$ group, target-dependent displacement of 0-40° were introduced, so that the reach direction cancelling the displacement was the same for all targets. This resulted in one reach direction being repeatedly rewarded. In the $Adp^{+}Rep^{-}$ the displacement was identical for all targets, so that participants adapted to a displacement but did not repeat a same rewarded reach direction. In the $Adp^{-}Rep^{+}$ group, participants reached to a single target with no displacement, so as to repeat a unique, rewarded reach direction in the absence of adaptation. In the $Adp^{-}Rep^{-}$, participants reached to 5 targets with veridical feedback. Before a re-learning block, the Adp^{+} performed reaches with veridical feedback to extinguish out the adaptation memory. The Adp^{-} groups transitioned directly to the re-learning block. B. During re-learning, participants reached to a single target with a displacement of 25° corresponding to the repeated reach direction. The $Adp^{+}Rep^{+}$ group showed faster re-learning compared to the other three groups. *Adp*: adaptation; *Rep*: repetition.

mechanism, whereby the memory of the hand solution that led to successful adaptation was strengthened because it was rewarded (*i.e.* it led to the cursor hitting the target); while the absence of repetition of the same hand solution in the control group prevented this reinforcement from occurring because it is a slow learning mechanism that requires repetition. Importantly, the authors argued toward a model-free form of reinforcement as opposed to model-based reinforcement, suggesting that it is implicit in nature (Haith & Krakauer, 2013; Otto et al., 2015).

A later study from the same group addressed the effect of reward on retention of a learnt visuomotor rotation (Shmuelof et al., 2012). In their task, participants successfully adapted to a 30° visuomotor displacement, before being exposed to a binary feedback that would inform them if they hit or miss the target. This binary feedback served as a rewarding signal, emphasizing the success/failure dimension of the task. In this study, they showed that introducing this binary feedback led to nearly complete retention of the visuomotor rotation, while introducing binary alongside visual feedback did not. In other words, participants in the binary-feedback-only group continued reaching to 30° off target even after all feedback had been removed, instead of decaying back to baseline.

Together, these studies advocated that the reward/failure dimension of the task can be employed to improve retention of a learnt visuomotor displacement. This result was later reproduced in a study looking at the differential effect of reward and punishment on motor adaptation (Galea et al., 2015). In this study, participants learnt a visuomotor displacement while being rewarded or punished with increasing amount of money as the cursor was close to target (reward group) and away from target (punish group), respectively. This led to a clear dissociation, where the rewarded group expressed higher retention values and the punished group expressed faster learning rates, therefore replicating previously seen effects of reward and demonstrating a new effect of punishment in the context of motor adaptation.

Altogether, those studies suggest the appealing possibility that learning and remembering can be manipulated, and critically enhanced, in motor adaptation if reward and punishment are used sensibly. This paves the way toward applications such as optimised rehabilitation procedures for clinical populations that have experienced a loss in motor ability or improving the training and performance of elite athletes and performers. Several studies have focused on assessing this possibility in the more applied context of rehabilitation, with promising results.

1.4.3 Rehabilitation and motor adaptation

A clinical population that may benefit from enhanced motor rehabilitation procedures is stroke patients, as they generally suffer from upper limb paresis. In an attempt to replicate the results from Galea et al. (2015), a study on stroke patients assessed adaptation performance of stroke patients in a force-field task with rewarding, punishing, or neutral feedback (Quattrocchi et al., 2017). Force-field paradigms require participants to adapt to viscous forces applied to their arm as a function of their velocity while they reach to a series of targets. Although it bears some difference with visuomotor adaptation, this paradigm leads to similar consequences in terms of motor adaptation. This study replicated most results from Galea et al. (2015), and further showed that both reward and punishment can lead not only to faster adaptation rates and stronger retention, but to higher final adaptation values compared to neutral groups in clinical populations. However, a limitation of this study is that albeit extending previous results to stoke populations, it merely tested individuals in lab-designed tasks and not in real-life rehabilitation procedures. Another study however assessed the effect of high versus low reward on stoke patients undergoing rehabilitation to restore ankle flexibility (Goodman et al., 2014). This study shows that not only did high reward feedback increase patients' flexibility to a much greater extent and with faster learning rates, but that their cortical efficiency (assessed using electro-encephalography) improved as well compared to the low-reward group.

Though the evidence confirming the positive effect of reward and punishment on plausible real-life rehabilitation procedures remains scarce, those studies provide a promising perspective. This motivates a need for better understanding the true essence of reinforcement in motor learning and especially in motor adaptation, which is the goal of this work.

1.5 Structure of the thesis

This work can be divided into two parts. Chapter 2 and chapter 3 focus on the role of reinforcement in motor adaptation, how it interacts with explicit control and how reward-based performance can be predicted for each individual. Chapter 4 and chapter 5 together represent the second part of this thesis. They are concerned with the impact of reward on motor control during reaching and how motor control can be enhanced by it. Finally, chapter 6 discusses the impact of this work on the literature and introduces possible future questions to be addressed.

Chapter 2

THE RELATIONSHIP BETWEEN REINFORCEMENT AND EXPLICIT MOTOR CONTROL DURING VISUOMOTOR ADAPTATION

This chapter has been accepted for publication in a peer-review journal as:
Codol O., Holland P.J. and Joseph M. Galea (2018). The relationship between reinforcement and explicit control during visuomotor adaptation. *Scientific Reports*.
It is also available as a pre-print on *bioRχiv* at the URL address `https://www.biorxiv.org/content/10.1101/206284v2`.

I helped design the tasks, implemented the tasks, acquired the data, analysed the data, interpreted the data, and wrote the chapter. P.J. Holland helped design the tasks and helped interpret the data. J.M. Galea helped design the tasks, helped interpret the data and provided the funding, facilities and materials to conduct the study.

2.1 Introduction

In a constantly changing environment, our ability to adjust motor commands in response to novel perturbations is a critical feature for maintaining accurate performance (Tseng et al., 2007). These adaptive processes have often been studied in the laboratory through the introduction of a visual displacement during reaching movements (Krakauer, 2009). The observed visuomotor adaptation, characterized by a reduction in performance errors, was believed to be primarily driven by a cerebellar-dependent process that gradually reduces the mismatch between the predicted and actual sensory outcome (sensory prediction error) of the reaching movement (Tseng et al., 2007; Wolpert & Miall, 1996; Wolpert et al., 1998). Cerebellar adaptation is a stereotypical, slow and implicit process and therefore does not require the individual to be aware of the perturbation to take place (Mazzoni & Krakauer, 2006; Shadmehr & Krakauer, 2008). However, a single-process framework cannot account for the great variety of results observed during visuomotor adaptation tasks (Taylor et al., 2014). Specifically, it has recently been shown that several other non-cerebellar learning mechanisms also play a pivotal role in shaping behaviour during adaptation paradigms such as explicit control (Taylor & Ivry, 2011, 2014) and reward-based reinforcement (Goodman et al., 2014; V. S. Huang et al., 2011; Izawa & Shadmehr, 2011; Kojima & Soetedjo, 2017; Quattrocchi et al., 2017; Shmuelof et al., 2012).

Explicit control usually consists of employing simple heuristics such as aiming off target in the direction opposite to a visual displacement, to quickly and accurately account for it (Mazzoni & Krakauer, 2006). However, this requires explicit knowledge of the perturbation, which in turn usually requires experiencing large and unexpected errors (Leow et al., 2016; Malfait, 2004; Orban de Xivry & Lefèvre, 2015; Taylor & Ivry, 2011). Explicit control contrasts with cerebellar adaptation in that it is idiosyncratic (Taylor & Ivry, 2014), volitional, and can lead to fast adaptation rates (Huberdeau et al., 2015). Importantly, in this work, we consider explicit control as the contribution to performance that can be suppressed (or expressed) by participants upon request (Werner et al., 2015), as opposed to the additional requirement of being able to verbalise a strategy. Critically, cerebellar adaptation takes place regardless of the presence or absence of any explicit process, even at the cost of accurate performance (Mazzoni & Krakauer, 2006).

More recently, another putative mechanism contributing to motor adaptation has been proposed, through which the memory of actions that led to successful outcomes (hitting the target) is strengthened, and therefore more likely to be re-expressed (Galea et al., 2015; Kojima & Soetedjo, 2017). Such reinforcement is considered to be an implicit process, but distinct from cerebellar adaptation in that it is not driven by sensory prediction error but task success or failure (V. S. Huang et al., 2011; Izawa & Shadmehr, 2011). To examine this phenomenon, several studies employed a hit-or-miss binary feedback paradigm which promotes reinforcement over cerebellar processes (Izawa & Shadmehr, 2011; Shmuelof et al., 2012; Therrien et al., 2016). For example, in one study, participants receiving only binary feedback following successful adaptation expressed stronger retention than participants who had received a combination of visual and binary feedback (Shmuelof et al., 2012). The authors argued this could be due to greater involvement of reinforcement-based process that is less susceptible to forgetting (Shmuelof et al., 2012).

With the multiple processes framework of motor adaptation, the question of interaction between the distinct systems becomes central to understanding the problem as a whole, and it remains an under-investigated question for reward-based reinforcement. In decision-making literature, it has long been suggested that two distinct "model-based" and "model-free" systems interact (Daw et al., 2011; Sun et al., 2005) and even require communication to be optimal (Gläscher et al., 2010; Huys et al., 2012). Interestingly, model-based processes share many characteristics with explicit control during motor adaptation, in that they are both more explicit, rely on an internal model of the world (explicit control: Haith & Krakauer, 2013; Hwang et al., 2006; model-based decision-making: Daw et al., 2005), and are closely related to working memory capacity (explicit control: Anguera et al., 2010; Christou et al., 2016; model-based decision-making: Otto, Gershman et al., 2013; Otto et al., 2015) and pre-frontal cortex processes (explicit control: Anguera et al., 2010; model-based decision-making: Gläscher et al., 2010; Simon & Daw, 2011. On the other hand, the concept of reinforcement in motor adaptation comes directly from the model-free systems described in decision-making literature (Haith & Krakauer, 2013), and is often labelled as such. It is considered more implicit, relies on immediate action-reward contingencies and is thought to recruit the basal ganglia in both cases (visuomotor adaptation: Therrien et al., 2016; decision-making: Daw et al., 2011). Despite these interesting similarities, unlike model-based and model-free decision-making, the relationship between explicit control and reinforcement during visuomotor adaptation paradigms is currently unknown. Evidence of this relationship exists from a recent study which showed participants needed to experience a large reaching error in order to express a reinforcement-based memory (Orban de Xivry & Lefèvre, 2015). In addition, there is a wealth of evidence which shows explicit control also requires experiencing large errors (Hwang et al., 2006; Leow et al., 2016; Malfait, 2004). Thus, it is possible that the formation of a reinforcement-based memory requires, or at least benefits, from some form of explicit control (Chen et al., 2017b).

To address this possibility, we first examined the contribution of explicit control to the reinforcement-based improvements in retention following binary feedback (Shmuelof et al., 2012; Therrien et al., 2016). We manipulated the amount of reinforcement participants were exposed to after adapting to a visuomotor displacement, and then tested how reinforcement altered retention of the motor memory in a subsequent block. To tease apart the explicit and implicit components of that memory, we asked some participants to "remove" any strategy they had and asked the others to "carry on as they were". Since an explicit contribution will be susceptible to volitional control by definition, participants who removed any strategy will only express the implicit component of a motor memory they formed, while those in the "carry on" group will express the combined contribution of the implicit and explicit components. The explicit contribution alone can therefore be inferred from the difference of the two. This design resulted in a 2×2 design with reinforcement versus no reinforcement and explicit retention versus no explicit retention. If explicit control is indeed important for enhancing reinforcement-based motor memories, an effect of the “remove” instruction should be observed, notably in the group that was exposed to reinforcement.

Secondly, we aimed at dissociating the explicit and implicit contribution of performance during exposition to reinforcement—as opposed to retention, which is tested after exposition to reinforcement. To that end, we used a forced fast reaction time (RT) paradigm which was shown to prevent expression of explicit control during a reaching task, and compared a condition with a slow RT in which participants can express explicit control (Haith et al., 2015). However, since that manipulation only allows to control for the expression of explicit control, we included a third, gradual condition that prevented development of an explicit strategy in the first place (Christou et al., 2016). If explicit control is indeed important to maintain performance when exposed to binary reinforcement feedback, the fast RT and gradual conditions should present altered reaching performance compared to the slow RT condition.

2.2 Methods

2.2.1 Participants

80 (20 male, mean age: 20.9, range: 18-37) and 30 (11 male, mean age: 22.1 years, range: 18-34) participants were recruited for experiment one and two, respectively, and pseudo-randomly assigned to a group after providing written informed consent. 14 additional participants were excluded from experiment one due to poor performance, in addition to the 80 participants whose dataset was included (see section 2.2.3 for details)—resulting in a total of 94 participants initially recruited. All participants were enrolled at the University of Birmingham. They were remunerated either with course credits or money

(£7.5/hour). They were free of psychological, cognitive, motor or auditory impairment and were right-handed. The study was approved by and done in accordance with the University of Birmingham Ethics Committee under the project code ERN_09-528P.

2.2.2 General procedure

Participants were seated before a horizontal mirror reflecting a screen above (refresh rate 60 Hz) that displayed the workspace and their hand position (figure 2.1A), represented by a green cursor (diameter 0.3cm). Hand position was tracked by a sensor taped on the right-hand index of each participant and connected to a Polhemus 3SPACE Fastrak tracking device (Colchester, Vermont U.S.A; sampling rate 120 Hz). Programs were run under Matlab (The Mathworks, Natwick, MA), with Psychophysics Toolbox 3 (Brainard, 1997). Participants performed the reaching task on a flat surface under the mirror, with the reflection of the screen matching the surface plane. All movements were hidden from the participant's sight. When each trial started, participants entered a white starting box (1cm width) on the centre of the workspace with the cursor, which triggered target appearance. Targets (diameter 0.5cm) were 8cm away from the starting position. Henceforth, the target position directly in front of the participant will be defined as the 0° position and other target positions will be expressed with this reference. Participants were instructed to perform a fast "swiping" movement through the target. Once they reached 8cm away from the starting box, the cursor disappeared and a yellow dot (diameter 0.3cm) indicated their end position. When returning to the starting box, a white circle displaying their radial distance appeared to guide them back.

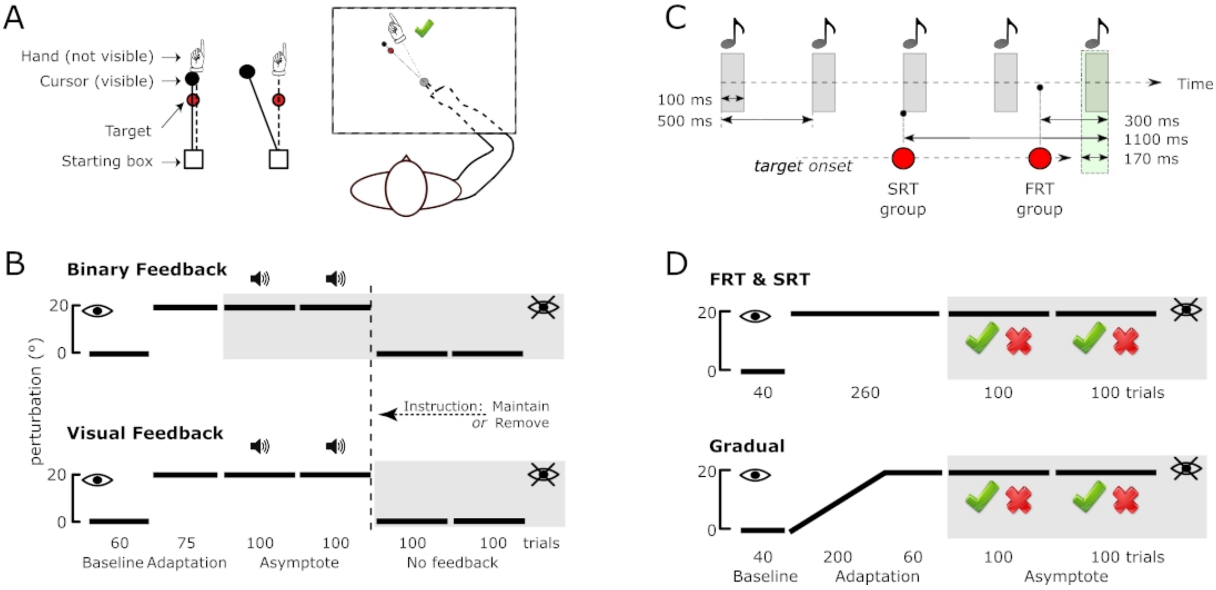


Figure 2.1: Experimental design. A. Experiment 1: feedback-instruction. Screen display and hand-cursor coupling before and after introduction of the visuomotor displacement (right and left, respectively). The rightmost part shows the experimental setup. B. Feedback-instruction task perturbation and feedback schedule for the BF (top) and VF groups (bottom). The white and grey areas represent blocks where visual feedback was available and not available, respectively, as indicated with a crossed or non-crossed eye. Blocks in which hits (±5° from target) were followed by a pleasant sound are indicated with a speaker symbol. The y-axis represents the discrepancy between hand movement and task feedback. The double dashed vertical lines represents the moment at which “Maintain” or “Remove” instructions were given. The block names and number of trials are indicated at the bottom of each schedule. C. Experiment 2: forced RT. Schedule of tone playback and target appearance before each trial for the SRT and FRT conditions. The green area represents the allowed movement initiation timeframe, and the red dots indicate the target onset for each condition. The grey areas represent the tones. D. Forced RT task perturbation and feedback schedule for the SRT, FRT (top) and Gradual groups (bottom). The nomenclature is the same as panel B. The green tick and red cross represent binary feedback cues for a hit (±5° from target) and miss, respectively. BF: binary feedback; VF: visual feedback; RT: reaction time; SRT: slow reaction time; FRT: fast reaction time.

2.2.3 Task design

Experiment 1: feedback-instruction

For each trial, participants reached to a target located 45° counter-clock wise (CCW). Participants first performed a baseline block (60 trials) with veridical cursor feedback, followed by a 75 trials adaptation block in which a 20° CCW displacement was applied (figure 2.1B). In the following 2 blocks (100 trials each), participants either experienced the same perturbation with only binary feedback, or with binary feedback and visual feedback. Binary feedback consisted of a pleasant sound selected based on each participant's preference from a series of 26 sounds before the task, unbeknownst of the final purpose. When participants' cursor reached less than 5° away from the centre of the target, the sound was played, indicating a hit; otherwise no sound was played, indicating a miss. For the binary feedback group (BF group), no cursor feedback was provided, except for one "refresher" trial every 10 trials where visual feedback was present. Participants in the visual feedback group (VF group) could see the cursor position during the outbound reach of the trial, along with the binary feedback. Finally, participants went through 2 no-feedback blocks (100 trials each) with binary and visual feedback completely removed. Before those blocks, participants were either told to "carry on" ("Maintain" group) or informed of the nature of the perturbation and asked to stop using any explicit approach to account for it ("Remove" group). Therefore, we had four groups in a 2×2 factorial design (BF versus VF and Maintain versus Remove).

It should be highlighted that the VF groups were in fact BF+VF groups since binary feedback was delivered alongside visual feedback. This choice was driven by two elements in the literature. First, the original article showing an effect of binary feedback on motor memories also showed that visual feedback in essence negates any effect of binary feedback—arguing as a possible explanation that visual feedback may be informative to the point that binary feedback information was ignored by the controller (Shmuelof et al., 2012). Second, visual feedback motor adaptation task is a very well-documented paradigm in the literature (Haith et al., 2015; Morehead et al., 2015; Galea et al., 2015, 2011; Kitago et al., 2013), including its explicit contribution (Taylor & Ivry, 2011, 2014; Bond & Taylor, 2015), and therefore, such a control would be less informative for this study. In addition to these two elements, BF+VF groups appeared as a closer, and therefore more valid control to the BF-only groups than VF-only groups. Therefore, although we acknowledge here that the VF groups are *de facto* VF+BF groups, we will refer to those groups as the VF groups for simplicity.

If a trial's reaching movement duration was greater than 400ms or less than 100ms long, the starting box turned red or green, respectively, to ensure participants performed ballistic movements, and didn't make anticipatory movements. Participants who expressed a success rate inferior to 40% during asymptote blocks were excluded, /textcolorredand recruitment continued until 20 successful participants per group was reached (participants removed: VF-Remove N=0; VF-Maintain N=0; BF-Remove N=6; BF-Maintain N=8). Although this exclusion rate was high, it was crucial to exclude participants who were unable to maintain asymptote performance in order to reliably measure retention.

Experiment 2: forced reaction times

In this experiment, participants (N=10 per group, 3 groups in total) were forced to perform the same reaching task at slow reaction time (SRT) or fast reaction time (FRT), the latter condition preventing explicit re-aiming by enforcing movement initiation before any mental rotation can be applied to the motor command (Fernandez-Ruiz et al., 2011; Haith et al., 2015). A third group (Gradual) also performed the task with no RT constraints.

In the SRT and FRT groups, for each trial, entering the starting box with the cursor triggered a series of five 100ms long pure tones (1 kHz) every 500ms (figure 2.1C). Before the fifth tone, a target appeared at one of four possible locations equally dispatched across a span of 360° (0-90-180-270°). Participants were instructed to initiate their movement exactly on the fifth tone (figure 2.1C). Targets appeared 1000ms (SRT) or 200ms (FRT) before the beginning of the fifth tone. Movement initiations shorter than 130ms are likely anticipatory movements (Haith et al., 2016), and explicit control starts to be difficult to express under 300ms (Haith et al., 2015; Leow et al., 2017). Therefore, in both conditions, movements were successful if participants exited the starting box between 70ms before the start of the fifth tone and the end of the fifth tone, that is, from 130ms to 300ms after target appearance in the FRT condition. If movements were initiated too early or too late, a message ”too fast” or ”too slow” was displayed and the cursor did not appear upon exiting the starting box. The trial was then reinitialised, and a new target selected. Finally, if participants repeatedly missed movement initiation, making trial duration over 25 seconds, RT constraints were removed, to allow trial completion before cerebellar memory time-dependent decay (S. Kim et al., 2015; Kitago et al., 2013; Yang & Lisberger, 2014). Participants in the SRT and FRT groups were informed of the displacement and of the optimal policy to counter it, to ensure that any effect was related to expression, rather than development of explicit control. They were also instructed to attempt using the optimal policy as much as possible when sensible, but not at the expense of the secondary RT task, so as to preserve the pace of the experiment and prevent time-dependent memory decay.

To attain proficiency in the RT task, SRT and FRT participants performed a training block (pseudo-random order of visual feedback and binary feedback trials) of at least 96 trials, or until they could initiate movements on the fifth tone reliably (at the first attempt) at least for 75% of the previous 8 trials. All participants achieved this in 96 to 157 trials. Once this was achieved, participants first performed a 40 trials baseline (figure 2.1D), followed by introduction of a 20° CCW displacement for 260 trials. Participants then underwent a 200-trials asymptote block with only binary feedback (1 “refresher” trial every 10 trials). The binary feedback consisted of a green tick or a red cross if participants hit or missed the target, respectively. Visual (instead of audio) binary feedback was used to avoid binary feedback sounds from lining up with the tones, which could potentially confuse participants. The Gradual group underwent the same schedule, except that no tone or RT constraint were used, and the perturbation was introduced gradually from the 41st to the 240th trial (increment of 0.4°/trial) occurring independently for each target. This ensured participants experienced as few large errors as possible to prevent awareness of the perturbation and therefore explicit control. After the experiment, participants in the Gradual group were informed of the displacement, and subsequently asked if they noticed it. If they answered positively, they were asked to estimate the size of the displacement.

2.2.4 Data analysis

All data and analysis code is available on our open science framework page (`osf.io/hrgzq`). All analyses were performed in Matlab. We used Lilliefors test to assess whether data were parametric, and we compared groups using Kruskal-Wallis or Wilcoxon signed-rank tests when appropriate, as most data were non-parametric. Post-hoc tests were done using Tukey's procedure. As we analysed the data from experiment two twice (figure 2.4C and 2.5), success rates and reach angles during asymptote were Bonferroni-corrected with corrected p-values (multiplied by 2).

Learning rates were obtained by fitting an exponential function to adaptation block reach angle curves with a non-linear least-square method and maximum 1000 iterations (average $R^2 = 0.86 \pm 0.14$ for feedback-instruction task and $R^2 = 0.58 \pm 0.26$ for forced-RT task):

$$y = a \cdot e^{\beta x} + b \tag{2.1}$$

where y is the hand direction for trial x, a is a scaling factor, b is the starting value and β is the learning rate. Reach angles were defined as angular error to target of the real hand position at the end of a movement. Trials were considered outliers and removed if movement duration was over 400ms or less than 100ms, end point reach angle was over 40° off target, and for the SRT and FRT groups in the forced-RT task, if failed initiation attempts continued for more than 25 seconds. In total, outliers accounted for 3755 trials (8%) in the feedback-instruction task and 1013 trials (6%) in the forced-RT task.

Even though 4 targets were used during the forced-RT task, trials were reset and a new random target was selected when participants failed to initiate movements on the 5th tone. Therefore, all possible target positions would not be represented for each epoch, and epochs were consequently not used.

2.3 Results

2.3.1 Experiment 1: explicit control drives performance during the recall of a reinforced motor memory

We first sought to investigate the role of explicit control in the retention of a reinforced visual displacement memory. In experiment 1, participants made fast 'shooting' movements towards a single target (figure 2.1A). After a baseline block involving veridical vision (60 trials) and an adaptation block (75 trials) where a 20° CCW visuomotor displacement was learnt with online visual feedback, participants experienced the same displacement for 2 blocks (asymptote blocks; 100 trials each) with either only binary feedback (BF group in figure 2.1B, top) to promote reinforcement, or binary feedback and visual feedback together (VF group in figure 2.1B, bottom). Following this, retention was assessed through 2 no-feedback blocks (100 trials each), during which both binary feedback and visual feedback were removed. Before these no-feedback blocks, half of the participants were told to "carry on" as they were ("Maintain" group) and the remaining ones were informed of the nature of the perturbation, and to stop re-aiming off target to account for it ("Remove" group). Thus, there were four groups: BF-Maintain, BF-Remove, VF-Maintain and VF-Remove (N=20 for each group).

Group performance is shown in figure 2.2A. All groups showed similar baseline performance (figure 2.2; $H(3) = 4.59, p = 0.20$) and had fully adapted to the visuomotor displacement prior to the asymptote/reinforcement blocks (average reach angle in the last 20 trials of adaptation, figure 2.2C; $H(3) = 2.56, p = 0.46$). Interestingly, at the start of the first asymptote block, participants in both BF groups showed a dip in performance, effectively drifting back toward baseline before adjusting back and returning to plateau performance. This "dip effect" was completely absent in the VF groups, and has previously been observed independently of our study when switching to binary feedback after a displacement is abruptly introduced (Shmuelof et al., 2012). Therefore, success rate

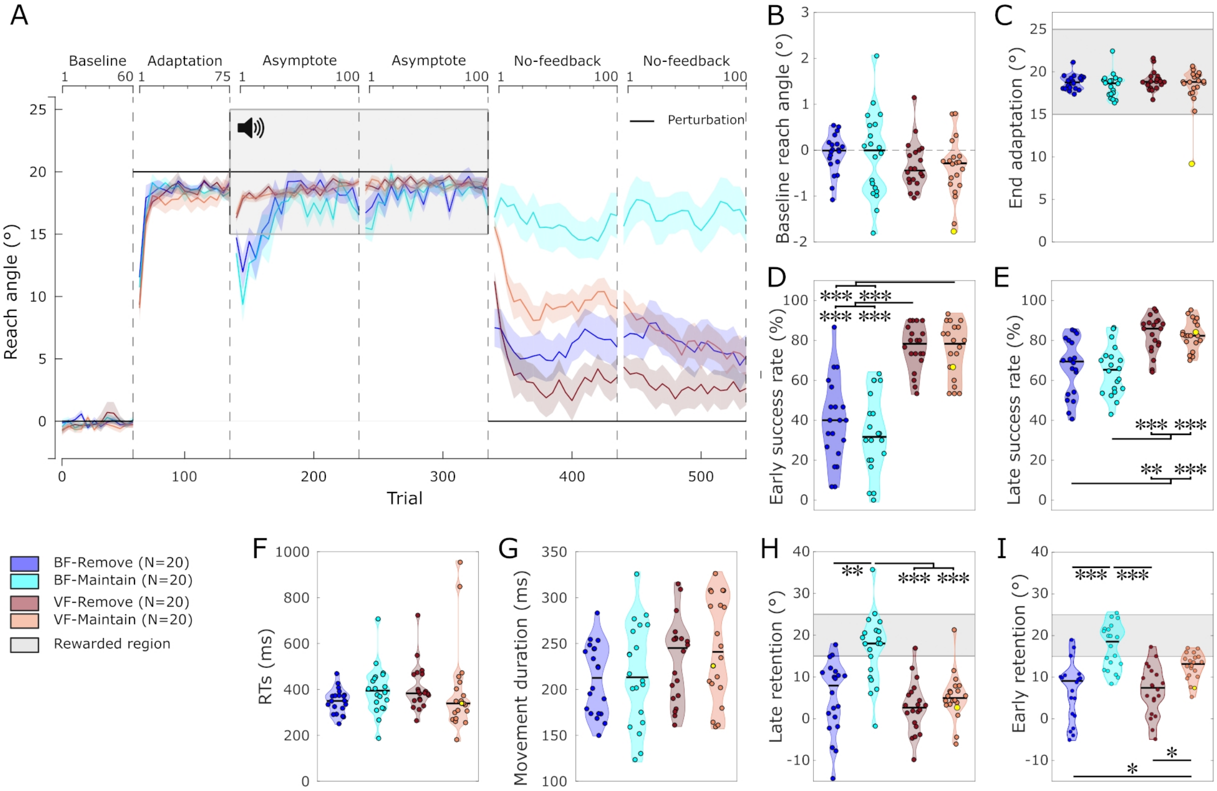


Figure 2.2: Experiment 1: feedback-instruction. A. Reach angles with respect to target of each group (N=20 per group) during the task. Values are averaged across epochs of 5 trials. Vertical bars represent block limits. The binary feedback consisted of a pleasant sound in the rewarded region. The black solid line represents the hand-to-cursor discrepancy (the perturbation) for all groups across the task. The upper and lower horizontal axes represent block-relative and absolute trial number, respectively. Coloured lines represent group mean and shaded areas represent s.e.m. B. Average reach angles during baseline. Of note, the y axis scale is smaller than in the following panels. C. Average reach angles in the last 20 trials of the adaptation block. The shaded area represents the region to be rewarded in the following block. D. Success rate (%) during the first 30 trials of the asymptote phase. E. Success rate during the remainder of the asymptote phase (trial 166-335). F. Average reaction times during the asymptote phase. G. Average movement durations during the asymptote phase. H. Average reach angle during the last 20 trials of the second no-feedback (retention) phase. I. Same as H but for the first 20 trials of the first no-feedback phase (early retention). Each dot represents one participant. The yellow dot represents the same participant across all plots, who expressed atypical end adaptation reach angles. For the distribution plots, horizontal black lines are group medians and the shaded areas indicate distribution of individual values. BF: binary feedback; VF: visual feedback. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

was compared independently across groups in the first 30 trials (figure 2.2D) and the remaining 170 trials (figure 2.2E) of the asymptote block. Both BF groups exhibited lower success rates than the VF groups in the early asymptote phase ($H(3) = 46.79, p < 0.001$, Tukey's test $p < 0.001$ for BF-Maintain vs VF-Maintain and vs VF-Remove, and for BF-Remove vs VF-Maintain and vs VF-Remove). This was also seen in the late asymptote phase ($H(3) = 31.29, p < 0.001$, Tukey's test $p < 0.001$ for BF-Maintain vs VF-Maintain and vs VF-Remove, and for BF-Remove vs VF-Maintain and vs VF-Remove), although performance greatly improved for both BF groups compared to the early phase ($Z = 3.692$ and $Z = -3.81$ for BF-Remove and BF-Maintain, respectively, $p < 0.001$ for both). Of note, both BF groups express a slight decrease in reach angle at the beginning of the second asymptote block but removing this second dip does not qualitatively alter the result ($H(3) = 27.46, p < 0.001$, Tukey's test $p < 0.001$ for BF-Maintain vs VF-Maintain and vs VF-Remove, and $p < 0.01$ for BF-Remove vs VF-Maintain and vs VF-Remove). Finally, no across-group difference in RTs or movement duration was found during the asymptote blocks (figure 2.2F, G).

Participants then performed a series of 2 no-feedback blocks. Similar to Shmuelof et al. (2012), we assessed retention by looking at the last 20 trials of the second block. However, our results are fundamentally the same irrespective of the trials used to represent late retention. Overall, the BF-Maintain group showed greater retention relative to all other groups, largely maintaining the reach angle values achieved during the asymptote phase, whereas there was no difference between the other groups (figure 2.2H; $H(3) = 27.66, p < 0.001$, Tukey's test $p = 0.001$ for BF-Remove vs BF-Maintain and $p < 0.001$ for BF-Maintain vs both VF groups; $p = 0.6$ for BF-Remove vs VF-Remove; $p = 1$ for BF-Remove vs VF-Maintain; $p = 0.68$ for VF-Maintain vs VF-Remove). We therefore replicated previous work which showed that binary feedback led to enhanced late retention of a visual displacement when compared to visual feedback (Shmuelof et al., 2012). However, this effect of binary feedback was abolished by asking participants to remove any re-aiming strategy they had developed (BF-remove). This suggests the increase in retention following binary feedback was mainly a consequence of the greater development and expression of explicit control.

However, based on the group data in figure 2.2A, it may be that early retention values lead to a different outcome. Therefore, we tested *a posteriori* the first 20 trials of the retention phase (figure 2.2I) for each group. Results show that both Maintain groups are different from the Remove groups ($H(3) = 32.27, p < 0.001$, Tukey's test $p < 0.001$ for BF-Maintain vs BF-Remove and BF-Maintain vs VF-Remove; $p = 0.02$ for VF-Maintain vs VF-Remove and $p = 0.03$ for VF-Maintain vs BF-Remove; $p = 0.99$ for BF-Remove vs VF-Remove; $p = 0.20$ for VF-Maintain vs BF-Maintain), leading to a clear dissociation between the Remove and Maintain instructions independently of feedback. This result suggests that what drives the difference between the Maintain and Remove condition in the visual feedback groups fades out gradually, while it does not in the binary feedback groups. Our interpretation is that this represents a cerebellar memory, that is, an implicit adaptation process that has been protected by the constant presence of visual feedback during the asymptote blocks. On the other hand, this cerebellar memory is likely to have extinguished during the asymptote blocks for the binary feedback groups due to the absence of visual feedback to feed it through sensory prediction errors. Therefore, this would suggest that what drives the residual difference in late retention between the BF-Maintain group and the other three is not cerebellar in nature, but rather the explicit component.

2.3.2 Experiment 2: re-aiming is necessary for maintaining performance under binary feedback

If the conclusion from our first experiment is correct, then successful asymptote performance under binary feedback only should be dependent on the ability to develop and express explicit control. Therefore, in experiment 2 we restricted participant's capacity to recruit an explicit component by using a forced RT adaptation paradigm (Haith et al., 2015, 2016; Leow et al., 2017; figure 2.1C and 2.3, see section 2.2 for details). Specifically, two groups adapted to a 20° CCW visuomotor displacement by performing reaching movements to 4 targets (figure 2.1D), with the amount of available preparation time (*i.e.* time between target appearance and movement onset) being restricted. A first group was allowed to express slow RTs. RT constraints were defined as 930 to 1100ms after target onset (N=10), while the second group was only allowed fast RTs (130 to 300ms; N=10; figure 2.1C). The latter condition has been shown to prevent time-demanding explicit processes such as mental rotations necessary to express re-aiming in reaching tasks (Fernandez-Ruiz et al., 2011; Haith et al., 2015; Leow et al., 2017). Critically, this paradigm prevented expression of re-aiming, but may not prevent development of an explicit component, at least reliably. Therefore, to ensure any between-group difference was task-dependent and not related to inter-individual differences in awareness or understanding of the task, we explained in detail the nature of the perturbation and the optimal policy to counter it. In addition, a third condition was designed in which participants were kept unaware of the visual displacement by introducing the perturbation gradually (Leow et al., 2016; Orban de Xivry & Lefèvre, 2015; N=10; figure 2.1D, bottom), and were not informed of any optimal policy to employ. Participants in this group were given no RT constraint whatsoever. Finally, it should be mentioned that a large portion of participants in the Gradual group reported noticing a slight perturbation by the end of the adaptation block when informally asked after the experiment. However, they underestimated its amplitude significantly at best, reporting effects of the order of 5°. Nevertheless, for the sake of simplicity we will qualify this group as "unaware", although we acknowledge they reported very partial, reduced awareness of the perturbation.

Overall group performance is displayed in figure 2.4A. During baseline, average reach direction was similar for all groups ($H(2) = 0.45, p = 0.79$; figure 2.4B). To examine whether the FRT and SRT groups displayed different rates of learning during adaptation, we applied an exponential model to each participant's adaptation data. Note, this was not done for the gradual group whose adaptation rate was restricted by the incremental visuomotor displacement. Surprisingly, we found no significant difference between the FRT and SRT group's learning rates ($U = 74; p = 0.34$; figure 2.4C). Indeed, one would

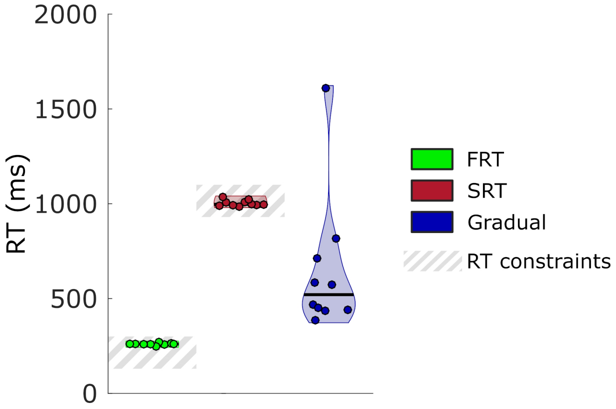


Figure 2.3: Reaction times expressed in the forced reaction time task. Individual dots indicate average reaction times of each participant throughout the task. For the distribution plots, horizontal black lines are group medians and the shaded areas indicate distribution of individual values. SRT: short reaction time; FRT: fast reaction time.

expect the SRT group to express faster learning since they can express strategies to account for the perturbation (Haith et al., 2015; Huberdeau et al., 2015; Leow et al., 2017; Morehead et al., 2015). This is most likely a consequence of the small size of the perturbation encountered (*i.e.* 20°), which leaves less margin for strategic re-aiming (Bond & Taylor, 2015; Morehead et al., 2015; Werner et al., 2015). At the end of the adaptation block, all groups adapted successfully, with no significant difference in reaching direction ($H(2) = 2.34, p = 0.31$; figure 2.4D). However, despite the lack of statistical significance, the mean reach direction for the FRT group was slightly under 15° (mean: 14.87°), which represents the limit of the reward region in the subsequent block. We discuss the implications of this later.

Participants then experienced an asymptote block with binary feedback, similar to the first experiment, with the exception that hit-miss feedback was provided with a green tick and a red cross onscreen, because audio binary feedback would potentially temporally align with movement initiation cues and confuse participants. Several other studies have already employed visual binary feedback successfully (Holland et al., 2018; Izawa & Shadmehr, 2011; Therrien et al., 2016). During asymptotic performance, where participants were restricted to binary feedback, the SRT group showed a striking ability to maintain

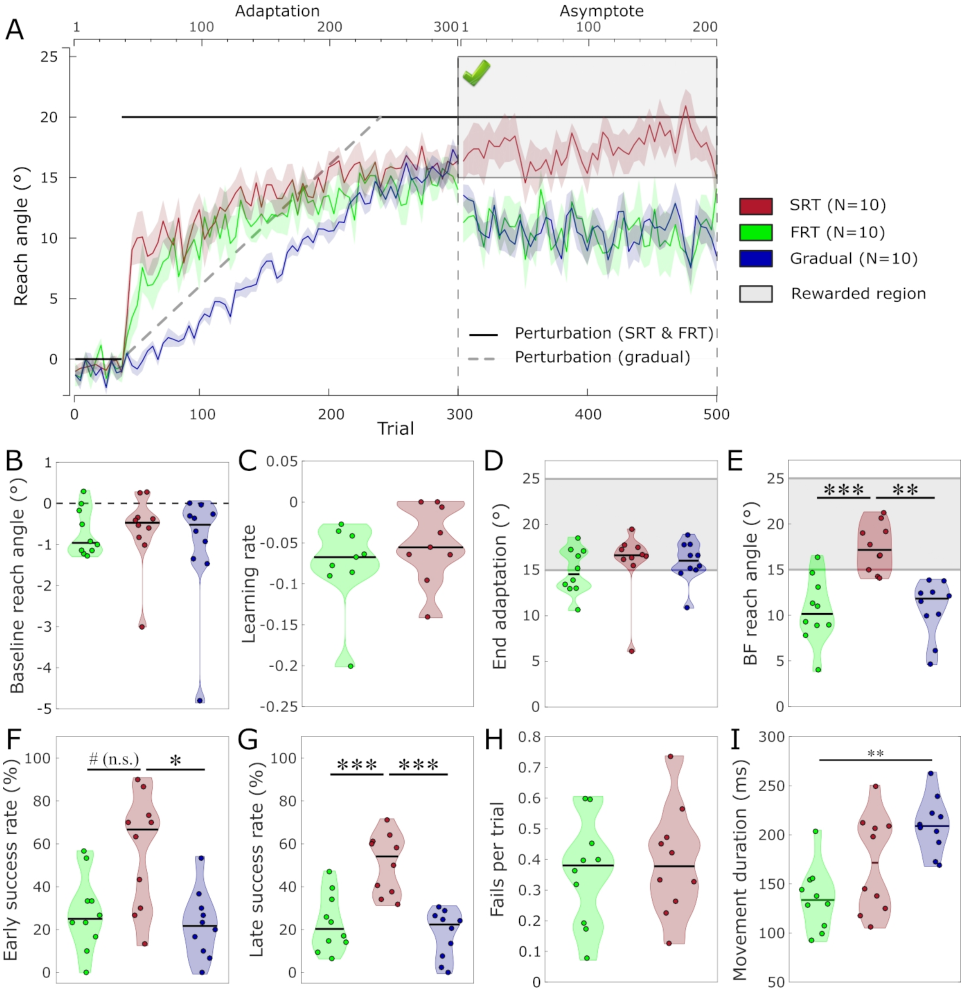


Figure 2.4: forced reaction times. Reach angles with respect to target of each group (N=10 per group). Values are averaged across epochs of 4 trials. Vertical bars represent block limits. The binary feedback consisted of a green tick displayed on top of the screen if participants were within the reward region (see figure), and of a red cross if not (not shown). The solid black and dashed grey lines represent the hand-to-cursor discrepancy (the perturbation) for the SRT and FRT group and for the Gradual group, respectively. The upper and lower horizontal axes represent block-relative and absolute trial number, respectively. Coloured lines represent group mean and shaded areas represent s.e.m. B. Average reach angle during baseline. Of note, the y axis scale is smaller than in the following figures. C. Learning rates during the adaptation block. D. Average reach angle during the last 20 trials of the adaptation block. The grey area represents the region to be rewarded in the subsequent block. E. Average reach angle during the asymptote block. F. Success rate during the first 30 trials of the asymptote phase. G. Success rate during the remainder of the asymptote phase (trial 331-500). H. Average number of failures per trial to initiate movements within the time constraints. I. Average movement duration. Each dot represents one participant. For the distribution plots, horizontal black lines are group medians and the shaded areas indicate distribution of individual values. SRT: short reaction time; FRT: fast reaction time. $\#p = 0.059$; *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

performance within the rewarded region whereas the two other groups clearly could not ($H(2) = 17.5, p < 0.001$, Bonferroni-corrected; figure 2.4E), Tukey's test $p < 0.001$ vs FRT and $p = 0.001$ vs Gradual). Next we compared success rates across groups for early binary feedback trials (figure 2.4F) and the remainder of binary feedback trials (figure 2.4G) independently. Early success rates were significantly lower for the Gradual group compared to the SRT ($H(2) = 9.2, p = 0.02$, Bonferroni-corrected, Tukey's test $p = 0.011$), and a similar but non-significant trend was observed between the FRT and SRT groups (Tukey's test $p = 0.059$). The absence of a significant difference in early success rate between the FRT and SRT groups cannot be explained by average reach angles, as the FRT group actually express a larger decrease in reach angle during that timeframe compared to the Gradual group (figure 2.4A). Rather, the greater variability in reach angle within individuals in the FRT as opposed to the Gradual group is likely to cause this result (average individual variance; FRT: 47.5; Gradual: 18.9). However, success rate during the remaining trials reached significance for both the FRT and Gradual groups compared to the SRT group ($H(2) = 16.67, p < 0.001$, Bonferroni-corrected, Tukey's test $p < 0.001$ for both FRT and Gradual). Surprisingly, no dip in performance was observed for the SRT group in the early phase of the binary feedback blocks, suggesting that informing participants of the perturbation and how to overcome it at the beginning of the experiment is sufficient to prevent this drop in reach angle.

Next, to ensure the low end adaptation reach angles expressed by the FRT group did not explain the low success rates, we removed every participant who expressed less than 15° reach angle at the end of the adaptation from each group (*e.g.* Saijo & Gomi, 2010). Henceforth, we refer to those participants as non-adapters, as opposed to adapters. This procedure resulted in 1, 5 and 2 participants being removed in the SRT, FRT and Gradual groups, respectively. Performance for the adapters was fundamentally the same as the original groups (figure 2.5A), except for end adaptation reach angles, which were now all above 15° (SRT 17.0 ±1.2; FRT 16.9 ±1.2; Gradual 16.7 ±1.4; figure 2.5B). Specifically, the SRT-adapter group still showed a clear ability to remain in the rewarded region during

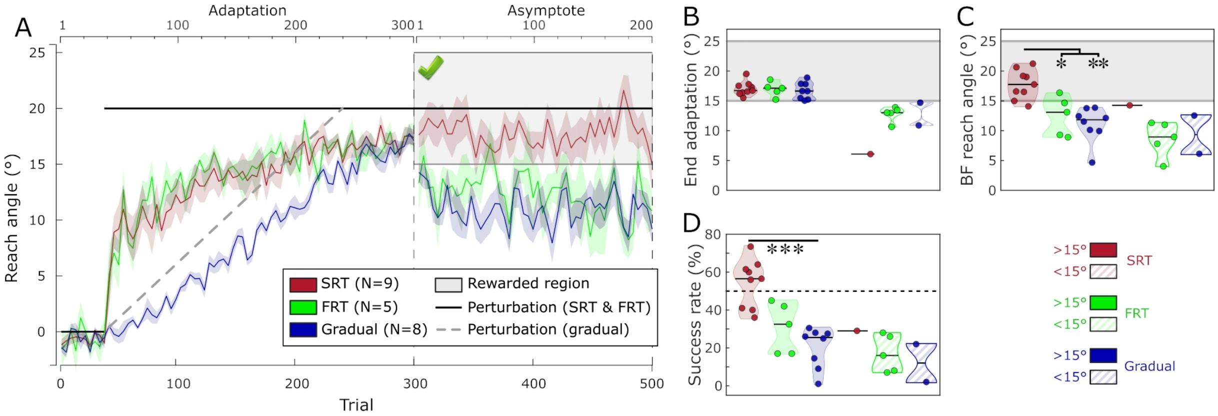


Figure 2.5: Performance of successful adapters during the forced reaction time task. A. Reach angles with respect to target of each group's successful adapters exclusively. Values are averaged across epochs of 4 trials. Vertical bars represent block limits. The binary feedback consisted of a large green tick displayed on top of the screen if participants were within the reward region (see figure), and of a red cross if they were not (not shown). The black solid line represents the hand-to-cursor discrepancy (the perturbation) for the SRT and FRT group across the task, and the grey dashed line represents the perturbation for the Gradual group only. The upper and lower horizontal axes represent block-relative and absolute trial number, respectively. Coloured lines represent group mean and shaded areas represent s.e.m. B. Average reach angle during the last 20 trials of the adaptation phase. The shaded area represents the region to be rewarded in the subsequent asymptote phase. C. Average reach angle during the binary feedback block. D. Success rate during the asymptote phase. The black dashed line represents 50% success rate. Each dot represents one participant. For the distribution plots, horizontal black lines are group medians and the shaded areas indicate distribution of individual values. >15° and <15° indicate the average reach angle during the end of the adaptation phase (*i.e.* adapter and non-adapter, respectively). SRT: short reaction time; FRT: fast reaction time. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

binary feedback performance (asymptotic blocks), whereas the other two adapter groups could not ($H(2) = 14.0, p = 0.002$, Bonferroni-corrected, Tukey's test $p = 0.028$ vs FRT-adapter and $p = 0.001$ vs Gradual-adapter; figure 2.5C). Because the full groups (*i.e.* non-Adapters included) did not express a drop in success rate during early asymptote trials, we compared Adapters' success rates during asymptote as a whole, rather than splitting them between early and late performance. The SRT-adapter group still displayed greater success than the Gradual-adapter group ($H(2) = 13.74, p = 0.002$, Bonferroni-corrected, Tukey's test $p < 0.001$; figure 2.5D). However, the difference between the SRT-adapter and the FRT-adapter group was now non-significant (Tukey's test $p = 0.12$). Despite this, the reach angle differences clearly show that successful binary performance remained strongly affected by one's capacity to develop and express explicit control even for the successful adapters, as shown by the Gradual-adapter and FRT-adapter groups, respectively (figure 2.5A).

Finally, since trials were reinitialised if participants failed to initiate reaching movements within the allowed timeframe, we compared the average occurrence of these failed trials between the FRT and SRT groups (figure 2.4H) to ensure any between-group difference cannot be explained by this. Both groups expressed similar amounts of failed attempts per trial ($U = 100, p = 0.73$). In addition, movement times were significantly faster across all blocks for the FRT group compared to the SRT group ($H(2) = 11.78, p = 0.005$, Tukey's test $p = 0.002$; figure 2.4I), although they remained strictly under 400ms for all groups as in the first experiment (figure 2.1C). This difference is to be expected due to the tendency to express faster velocities in movements with rapid initiation (Orban de Xivry et al., 2017). RTs expressed by the Gradual group were between the SRT and FRT constraints (figure 2.3; Gradual group RT range: 385 to 1610ms).

Overall these findings demonstrate that preventing explicit control by restricting its expression or making participants unaware of the nature of the task results in the partial incapacity of participants to perform successfully during binary feedback performance. It should be noted, however, that performance did not reduce back to baseline entirely, as participants in both the FRT and Gradual groups were still able to express intermediate reach angle values in the order of 10 to 15°.

2.4 Discussion

Previous work has led to the idea that binary feedback induces the recruitment of a model-free reinforcement system that strengthens and consolidates the acquired memory of a visuomotor displacement (V. S. Huang et al., 2011; Shmuelof et al., 2012; Therrien et al., 2016). Here, we investigated the role of explicit control in the context of binary feedback, and our results suggest that it may have a more central role in explaining some binary-feedback induced behaviours than previously expected. In the first experiment, the increased retention observed in the BF-Maintain group was suppressed if participants were told to "stop aiming off target" (BF-Remove group). In the second experiment, preventing expression of explicit control by using a secondary task or preventing its development with a gradual introduction of the perturbation resulted in participants being unable to maintain accurate performance during binary feedback blocks. This suggests an explicit component is necessary for performing a binary feedback reaching task, at least within the present study's experimental design.

The initial performance drop observed at the introduction of binary feedback for both BF groups suggests that participants cannot immediately account for a visuomotor displacement they have already successfully adapted to (Shmuelof et al., 2012). A possible explanation is that the cerebellar memory is not available anymore, most likely because removing visual feedback results in a context change, which is known to prevent retrieval and expression of an otherwise available memory (Brennan & Smith, 2015; Pekny et al., 2011; Smith et al., 2006). Considering this, the restoration of performance observed after this dip could not be explained by recollection of the cerebellar memory, suggesting another mechanism took place. Two possible candidates to explain this drift back are model-free reinforcement (V. S. Huang et al., 2011; Izawa & Shadmehr, 2011; Shmuelof et al., 2012; Therrien et al., 2016) and explicit processes (Bond & Taylor, 2015; Taylor et al., 2014; Taylor & Ivry, 2011).

Reinforcement learning is usually considered to operate through experiencing success (Chen, Holland & Galea, 2018; V. S. Huang et al., 2011; Izawa & Shadmehr, 2011). It is thus difficult to argue for a reinforcement-based reversion to good performance during binary feedback because participants in the trough of the dip did not experience a large amount of success, if any. Furthermore, participants experienced little "plateau" performance during the previous block, making formation of a model-free reinforcement memory unlikely, because it is considered a rather slow learning process as opposed to model-based reinforcement (V. S. Huang et al., 2011; Sutton & Barto, 1998); though the adaptation block remains longer compared to Shmuelof et al. (2012). On the other hand, both BF groups experienced a large amount of unexpected errors during this drop, which may promote a more explicit approach (Chen et al., 2017b; Hwang et al., 2006; Leow et al., 2016; Malfait, 2004; Orban de Xivry & Lefèvre, 2015). In line with this, the SRT group in the forced RT task, which had been informed of the displacement and of the right policy to counter it, did not express such a dip when starting the binary feedback block.

The forced RT task addresses this question more directly, and shows that impeding explicit control with a secondary task (Haith et al., 2015; Leow et al., 2017) prevents participants from restoring performance over binary feedback blocks, confirming our interpretation. Interestingly, both the FRT and Gradual groups did not show a return to baseline during asymptote. Likely, the FRT group was aware of the optimal policy, and could partially express it, leading to these intermediate reach angles. In line with this, previous work on forced RT paradigms shows that adapting the constraints based on each individual's baseline proficiency at this task more efficiently prevents explicit control (Leow et al., 2017). Furthermore, even in the presence of binary feedback, the Gradual group showed a striking inability to find the optimal policy, suggesting the lack of structural understanding of the task strongly impeded their exploration (Chen et al., 2017b; Chen, Holland & Galea, 2018). This overall incapacity of the Gradual group to express an efficient explorative approach is consistent with previous findings showing that rewarding success alone, without providing any explanation of the task structure, is not sufficient to make participants reliably learn an optimal policy (Chen, Holland & Galea, 2018; Manley et al., 2014).

However, an alternative explanatation that could be tested in the future would be that participants in the FRT and Gradual groups would eventually show an improvement in performance during the asymptote blocks if they were exposed to a sufficient amount of rewarding feedback. One way of implementing this would be to employ a closed-loop feedback design, whereby the amount of reward provided is not a function of absolute but relative performance of the participant (Therrien et al., 2016).

Previous studies employing the forced RT paradigm have shown it usually leads to slower learning rates during adaptation because participants can less easily employ explicit control from the beginning (Haith et al., 2015; Huberdeau et al., 2015; Leow et al., 2017). In contrast, no such difference in learning rate was observed in our forced RT groups. This is possibly due to the difference in size of the perturbation between our study (20°) compared to others (Haith et al., 2015; Leow et al., 2017) (30°), making the explicit contribution potentially smaller during the adaptation phase (Taylor et al., 2014).

Our findings qualitatively replicate results from a previous study employing a similar design (Shmuelof et al., 2012). However, it should be noted that our paradigm differs in several ways. First, retention was assessed using feedback removal rather than visual error clamps, although there is evidence that both methods lead to quantitatively similar results (Kitago et al., 2013). Second, our displacement was only 20° of amplitude and no additional displacement was introduced after the asymptote blocks. There is now a growing wealth of evidence that the cerebellum cannot account for more than 15 to 20° displacements (Leow et al., 2017; Morehead et al., 2017; Werner et al., 2015), with the remaining discrepancy usually being accounted for through explicit re-aiming (Bond & Taylor, 2015). Therefore, the absence of a second, larger displacement, if anything, should only result in a less explicit performance. Nevertheless, instructing participants to remove any explicit re-aiming policy (Remove groups) resulted in a near-complete nullification of the binary feedback effect, suggesting it is mainly underlain by a simple re-aiming process. However, the Maintain instruction alone was not sufficient to produce this high retention profile, as the VF-Maintain group did not express it. We believe this can be explained in two ways. First, experiencing no feedback may result in a stronger context change for the VF groups compared to the BF groups, because the latter experienced the absence of visual feedback during the asymptote blocks beforehand. Thus, this should lead to a stronger drop in reaching angle at the beginning of the no feedback trials for the VF groups, as observed here. Alternatively, the VF-Maintain group experienced 200 more trials with visual feedback at asymptote. Consequently, it is very likely that the cerebellar memory at the beginning of the no-feedback blocks was stronger (Izawa & Shadmehr, 2011), and the explicit contribution was less for this group compared to the BF-Maintain group (Bond & Taylor, 2015; Huberdeau et al., 2015; McDougle et al., 2015; Taylor et al., 2014). This would therefore result in the slow drop in reach angle observed during early no-feedback trials due to gradual decay of the cerebellar memory (Brennan & Smith, 2015; Kitago et al., 2013; Yang & Lisberger, 2014). Critically, both possibilities are not incompatible, and may well occur together.

A notable feature of retention performance is that both BF- and VF-Remove groups show a residual bias of around 5° in their reach angle in the direction opposite to the displacement. Participants in the Remove conditions were not aware of this upon asking them after the experiment. This has been reliably observed in studies using no-feedback blocks to assess retention (Galea et al., 2011, 2015) (but see Kitago et al., 2013). Possible explanations include use-dependent plasticity-induced bias (Bütefisch et al., 2000; Classen et al., 1998), perceptual bias (Vindras et al., 1998) or an implicit model-free reinforcement-based memory, although this study cannot provide any account toward one or the other. Note however that although the BF-Remove group expressed slightly more bias than its visual feedback counterpart, this clearly did not reach statistical significance, meaning this cannot be explained by feedback type alone. Regardless, the implicit and lasting nature of this phenomenon makes it a promising focus for future research with clinical applications (Goodman et al., 2014; Quattrocchi et al., 2017).

Overall, our findings point towards a central role of explicit control during binary-feedback induced behaviours in this study. In line with this, 14/54 participants had to be removed from the BF groups in the feedback-instruction task (experiment 1) because of poor performance in the asymptote blocks (see methods), suggesting that structural learning was required to perform accurately (Chen, Holland & Galea, 2018; Chen et al., 2017b; Manley et al., 2014). Though this is a significant proportion of participants, it should be noted that other studies on reaching under binary feedback also found a similar percentage of "learners" and "non-learners" (Holland et al., 2018; Saijo & Gomi, 2010). Although not expected in our study, this seemingly consistent outcome across a variety of binary feedback experimental designs raises questions regarding either the reliability of this learning mechanism across individuals or the tasks used to examine it. The possibility that this dichotomy between participants is due to structural learning is in line with the dip observed in the BF groups and the absence of dip in the (*i.e.* informed) SRT group. If correct, then predictors of structural learning capacity should also predict an individual's ability to learn a visuomotor displacement under binary feedback, a hypothesis that will be tested in future studies. Finally, our view is that implicit, model-free reinforcement takes a great amount of time and practice to form (Sutton & Barto, 1998; Wunderlich et al., 2012), and usually arises from initially model-based performance in behavioural literature (Daw et al., 2011; Tricomi et al., 2009), as illustrated by popular reinforcement models (*e.g.* DYNA; Sutton, 1990; Sutton et al., 2008). Two interesting possibilities are that 200 trials of binary feedback alone are not sufficient to result in a strong, habit-like enhancement of retention (Tricomi et al., 2009), or that such behavioural consolidation must take place through sleep (Reis et al., 2009; Tricomi et al., 2009). Future work is required to address these hypotheses.

In conclusion, this study provides further insight into the use of reinforcement during motor learning, and suggests that successful reinforcement is tightly coupled to the development and expression of explicit control. We suggest that explicit control bears many similarities with model-based reinforcement, thus creating important questions regarding the link between model-based and model-free reinforcement systems during motor learning. At the very least, future studies investigating reinforcement during visuomotor adaptation should proceed with care in order to map which behaviour is the consequence of implicitly reinforced memories or explicit control.

Chapter 3

DOMAIN-SPECIFIC WORKING MEMORY, BUT NOT DOPAMINE-RELATED GENETIC VARIABILITY, SHAPES REWARD-BASED MOTOR LEARNING

This chapter describes a study of which I have been one of the main contributors. A current draft is available on *bioRχiv* at the URL address `https://www.biorxiv.org/content/10.1101/524900v2`. Holland P.J.‡, Codol O.‡, Oxley E., Taylor M., Hamshere E., Joseph S., Huffer L., Galea M.J. (2019). Domain-specific working memory, but not dopamine-related genetic variability, shapes reward-based motor learning. *bioRχiv*.
‡ equal contribution.

This chapter is based on two tasks referred to as the Acquire and the Preserve task. The Acquire task was implemented and analysed by P.J. Holland, and he acquired most of the dataset, with additional help from E. Oxley, M. Taylor and E. Hamshere. The Preserve task was implemented and analysed by me, and I acquired most of the data, with additional help from S. Joseph and L. Huffer. The mediation analysis was performed by P.J. Holland and a lasso regression analysis that was eventually not included was performed by me. Conceptualisation of the study, task design and interpretation of results were done conjointly by J.M. Galea, P.J. Holland and me. J.M. Galea provided the fundings and materials. I wrote the first draft of the pre-print which serves as the basis for this chapter, and the final version was approved by every author.

Note: because the previous chapter was published before this work was written, it is occasionally referred to as Codol et al. (2018) instead of chapter 2.

3.1 Introduction

When performing motor tasks under altered environmental conditions, adaptation to the new constraints occurs through the recruitment of a variety of systems (Taylor & Ivry, 2014). Arguably the most studied of those systems is cerebellum-dependent adaptation, which consists of the implicit and automatic recalibration of mappings between actual and expected outcomes, through sensory prediction errors (Morehead et al., 2017; Tseng et al., 2007). Besides cerebellar adaptation, other work has demonstrated the involvement of a more cognitive, deliberative process whereby motor plans are adjusted based on an individual's structural understanding of the task (Bond & Taylor, 2015; Taylor & Ivry, 2011). We label this process "explicit control" (Codol et al., 2018; Holland et al., 2018) but it has also been referred to as strategy (Taylor & Ivry, 2011) or explicit re-aiming (Morehead et al., 2015). Recently it has been proposed that reinforcement learning, whereby the memory of successful or unsuccessful actions are strengthened or weakened, respectively, may also play a role (V. S. Huang et al., 2011; Izawa & Shadmehr, 2011; Shmuelof et al., 2012). Such reward-based reinforcement has been assumed to be an implicit and automatic process (Haith & Krakauer, 2013). However, recent evidence suggests that phenomena attributed to reinforcement-based learning during visuomotor rotation tasks can largely be explained through explicit processes (see chapter 2; Holland et al., 2018.

One outstanding feature of reinforcement-based motor learning is the great variability expressed across individuals (Codol et al., 2018; Holland et al., 2018; Therrien et al., 2016, 2018). What factors underlie such variability is unclear. If reinforcement is indeed explicitly grounded, it could be argued that individual working memory capacity (WMC), which reliably predicts propensity to employ explicit control in classical motor adaptation tasks (Anguera et al., 2010; Christou et al., 2016; Sidarta et al., 2018), would also predict performance in a reinforcement-based motor learning task. If so, this would strengthen the proposal that reward based motor learning bears a strong explicit component. Anguera et al. (2010) demonstrated that mental rotation working memory (RWM), unlike other forms of working memory such as verbal working memory (VWM), correlates with explicit control. More recently, Christou et al. (2016) reported a similar correlation with spatial working memory (SWM).

Another potential contributor to this variability is genetic profile. In chapter 2 and previous work (Holland et al., 2018), we argue that reinforcement-based motor learning performance relies on a balance between exploration and exploitation of the task space, a feature reminiscent of structural learning and reinforcement-based decision-making (Daw et al., 2005; Frank et al., 2009; Sutton & Barto, 1998). A series of studies from Frank and colleagues suggests that individual tendencies to express explorative/exploitative behaviour can be predicted based on dopamine-related genetic profile (Doll et al., 2016; Frank et al., 2009, 2007). Reinforcement has consistently been linked to dopaminergic function in a variety of paradigms, and thus, such a relationship could also be expected in reward-based motor learning (Pekny et al., 2015). Specifically, Frank and colleagues focused on Catecholamine-O-Methyl-Transferase (COMT), Dopamine- and cAMP-Regulated neuronal Phosphoprotein (DARPP32) and Dopamine Receptor D2 (DRD2), and suggest a distinction between COMT-modulated exploration and DARPP32- and DRD2-modulated exploitation (Frank et al., 2009).

Consequently, we investigated the influence of WMC (RWM, SWM, and VWM) and genetic variations in dopamine metabolism (DRD2, DARP32, and COMT) on an individual's ability to perform reward-based motor learning. A relationship with WMC would suggest that reinforcement-based motor learning bears an explicit component, in line with previous reports (Codol et al., 2018; Holland et al., 2018). In addition, a relationship with dopamine-related genes would provide original evidence of genetic influence in these tasks. Importantly, both those possibilities may be observed conjointly, as they are not mutually exclusive. If a relationship is observed, we expect that higher WMC will lead to faster learning, and possibly to a larger explicit component in motor memory. Regarding genetic profiles, allelic variations that result in higher dopaminergic metabolism should also lead to faster learning, but possibly also to higher implicit—as opposed to explicit—memory. It should be noted that beyond these expectations, we did not make predictions regarding which of our predictors will yield a significant relationship, as this aspect of this chapter remains exploratory.

We tested our predictors using two established reward-based motor learning tasks. First, a task analogous to a gradually introduced rotation (Holland et al., 2018) required participants to learn to adjust the angle at which they reached to a slowly and secretly shifting reward region signalled by binary feedback. Since reward was introduced from the learning phase itself, this task allowed to observe how participants could acquire a new reaching direction via rewarding feedback alone, and so we labelled it *Acquire* task. At the end of the task, the motor memory of the task was tested using a first *maintain* block in which participants were instructed to *"carry on as they were"*, allowing to observe the combined explicit and implicit components of the motor memory, similar to the feedback-instruction task in chapter 2. Following this maintain block, a *remove* block in which participants were told to *"Remove any strategy they had and start aiming directly toward the target"* allowed to subsequently assess the implicit component of the memory. In addition to the Acquire task, a task with an abruptly introduced rotation (Codol et al., 2018; Shmuelof et al., 2012) required participants to preserve performance with reward-based feedback after adapting to a visuomotor rotation. Here, since participants acquired the new reaching direction before the reward-based feedback was introduced, the task allowed to observe the participants' capacity to preserve a previously acquired motor performance while experiencing a change in feedback from visual- to reward-based. Accordingly, this task was labelled *Preserve* task. Therefore, the use of these two tasks enabled us to examine whether similar predictors of performance explained participant's capacity to acquire and preserve behaviour with reward-based feedback.

3.2 Methods

Prior to the start of data collection, the sample size, variables of interest and analysis method were pre-registered based on previous studies. The pre-registered information, data and analysis code can be found online at on the Open Science Framework website at `https://osf.io/j5v2s/` for the Perserve task and at `https://osf.io/rmwc2/` for the Acquire tasks.

3.2.1 Participants

121 (30 male, mean age: 21.06, range: 18-32) and 120 (16 male, mean age: 19.24, range: 18-32) participants were recruited for the Acquire and Preserve tasks, respectively. All participants provided informed consent and were remunerated with either course credit or money (£7.50/hour). All participants were free of psychological, cognitive, motor or uncorrected visual impairment. The study was approved by and done in accordance with the University of Birmingham Ethics Committee under the project code ERN_09-528P.

3.2.2 Experimental design

Participants were seated before a horizontally fixed mirror reflecting a screen placed above, on which visual stimuli were presented. This arrangement resulted in the stimuli appearing at the level on which participants performed their reaching movements. The Acquire (gradual) and Preserve (abrupt) tasks were performed on two different stations, with a KINARM (BKIN Technology, London, Ontario; sampling rate: 1000Hz) and a Polhemus 3SPACE Fastrak tracking device (Colchester, Vermont U.S.A.; sampling rate: 120Hz), employed respectively. The Acquire task was run using Simulink (The Mathworks, Natwick, MA) and Dexterit-E (BKIN Technology), while the Preserve task was run using Matlab (The Mathworks, Natwick, MA) and Psychophysics toolbox (Brainard, 1997). The Acquire task employed the same paradigm and equipment as Holland et al. (2018), with the exception of the maximum RT which was increased from 0.6s to 1s, and the maximum movement time (MT) which was reduced from 1s to 0.6s. The Preserve task used the same setup and display as in chapter 2; however, the number of 'refresher' trials during the binary feedback blocks was increased from one to two in every 10 trials because a pilot experiment showed that the majority of participants were failing at the Preserve task with only one refresher trial per 10 trials. The designs were kept as close as possible to their respective original publications to promote replication and comparability across studies. Consequently, all design parameters were taken from those previous studies unless specified otherwise. In both tasks, reaching movements were made with the dominant arm.

3.2.3 Reaching tasks

Acquire task

Participants performed 670 trials, each of which followed a stereotyped timeline. The starting position for each trial was in a consistent position 30cm in front of the midline and was indicated by a red circle (1cm radius). After holding the position of the handle within the starting position, a target (red circle, 1cm radius) appeared directly in front of the starting position at a distance of 10cm. Participants were instructed to make a rapid "shooting" movement that passed through the target. They experienced binary feedback similar to chapter 2: If the cursor position at a radial distance of 10cm was within a reward region (±5.67°, initially centred on the visible target; grey region in figure 3.1A) the target changed colour from red to green and a green tick was displayed just above the target position, informing participants of the success of their movement. If, however, the cursor did not pass through the reward region, the target disappeared from view and no tick was displayed, signalling failure. After each movement, the robot returned to the

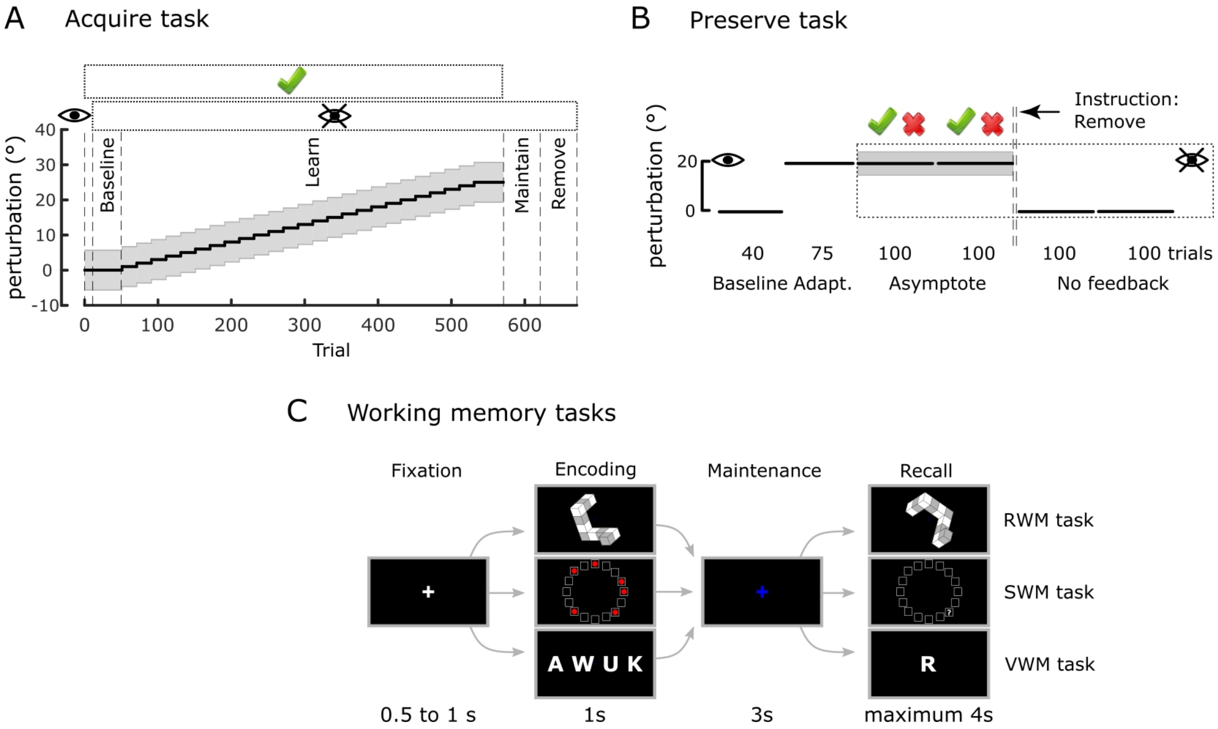


Figure 3.1: Experimental design. A. Time course of the Acquire task with the different experimental periods labelled. The grey region represents the reward region, which gradually rotated away from the visual target after the initial baseline period. The rectangle enclosing the green tick above the axes represents trials in which reward was available, and the rectangle with the "eye" symbol indicates when vision was not available. B. Time course of the Preserve task. After adapting to an initial rotation with vision available, vison was removed (eye symbol) and reward-based feedback was introduced (tick and cross above the axes). Prior to the no-feedback blocks participants were instructed to remove any strategy they had been using. C. WMC tasks, the three tasks followed a stereotyped timeline with only the items to be remembered differing. Each trial consisted of 4 phases (Fixation, Encoding, Maintenance, and Recall) with the time allocated to each displayed below, in seconds. WMC: working memory capacity; RWM: Mental rotation working memory; SWM: Spatial working memory; VWM: Verbal working memory.

starting position and participants were instructed to passively allow this.

For the first 10 trials, the position of the robotic handle was displayed as a white cursor (0.5cm radius) on screen, following this practice block the cursor was extinguished for the remainder of the experiment and participants only received binary feedback. The baseline block consisted of the first 40 trials under binary feedback, during this period the reward region remained centred on the visible target. Subsequently, unbeknownst to the participant the reward region rotated in steps of 1° every 20 trials; the direction of rotation was counterbalanced across participants. After reaching a rotation of 25° the reward region was held constant for an additional 20 trials. Performance during these last 20 trials was used to determine overall task success. Subsequently, binary feedback was removed, and participants were instructed to continue reaching as they were (maintain block) for the following 50 trials. Following this, participants were then informed that the reward region shifted during the experiment but not of the magnitude or the direction of the shift. They were then instructed to return to reaching in the same manner as they were at the start of the experiment (remove block, 50 trials). During the learning phase of the task participants were given a 1-minute rest after trials 190 and 340.

Preserve task

Participants performed 515 trials in total. On each trial participants were instructed to make a rapid "shooting" movement that passed through a target (white circle, radius: 0.125cm) visible on the screen. The starting position for each trial was indicated by a white square (width: 1cm) roughly 30cm in front of the midline and the target was located at angle of 45° from the perpendicular in a counter clockwise direction at a distance of 8cm. The position of the tracking device attached to the fingertip was displayed as a cursor (green circle, radius: 0.125cm). When the radial distance of the cursor from the starting position exceeded 8cm, the cursor feedback disappeared, and the end position was displayed instead.

First, participants performed a baseline period of 40 trials, during which the position of the cursor was visible and the cursor accurately reflected the position of the fingertip. In the adaptation block (75 trials), participants were exposed to an abruptly introduced 20° clockwise visuomotor rotation of the cursor feedback (figure 3.1B). Subsequently, all visual feedback of the cursor was removed, and participants received only binary feedback. If the end position of the movement fell within a reward region, the trial was considered successful and a tick was displayed; otherwise a cross was displayed. The reward region was centred at a clockwise rotation of 20° with respect to the visual target with a width of 10°, *i.e.* it was centred on the direction that successfully accounted for the previously experienced visuomotor rotation. Binary feedback was provided for 200 trials divided into 2 blocks of 100 trials (asymptote blocks). Furthermore, participants experienced 2 "refresher" trials for every 10 trials, where rotated visual feedback of the cursor position was again accessible (Codol et al., 2018; Shmuelof et al., 2012). This represents an increase compared to chapter 2 because participants in this study tended to have poorer performance under binary feedback, possibly due to a fatigue effect following the working memory (WM) tasks (Anguera et al., 2012; see discussion in section 3.4), though this last point is only conjecture. This was decided following a pilot study in which the majority of participants did not manage to obtain more than 40% success during the asymptote blocks. Finally, two blocks (100 trials each) with no performance feedback were employed in order to assess retention of the perturbation (no-feedback blocks). Before the first of those two blocks, participants were informed of the visuomotor rotation, asked to stop accounting for it through aiming off target and to aim straight at the target, similar to the binary feedback-remove group in the feedback-instruction experiment in chapter 2.

3.2.4 Working memory tasks

Participants performed three WM tasks, all of which followed the same design with the exception of the nature of the items to be remembered (figure 3.1C). All WM tasks were run using Matlab (The Mathworks, Natwick, MA) and Psychophysics toolbox (Brainard, 1997). At the start of each trial, a white fixation cross was displayed in the centre of the screen for a period of 0.5 to 1s randomly generated from a uniform distribution (fixation period; figure 3.1C). In the encoding period, the stimuli to be remembered was displayed for 1s and then subsequently replaced with a blue fixation cross for the maintenance period which persisted for 3s. Finally, during the recall period, participants were given a maximum of 4s to respond by pressing one of three keys on a keyboard with their dominant hand. The “1” key indicated that the stimuli presented in the recall period was a “match” to that presented in the encoding period, the “2” key indicated a “non-match” and pressing “3” indicated that the participant was unsure as to the correct answer. Each WM task contained 5 levels of difficulty with the 12 trials presented for each; 6 of which were trials in which “match” was the correct answer and 6 in which “non-match” was the correct answer. Consequently, each task consisted of 60 trials and the order in which the tasks were performed was pseudo-randomised across participants. Prior to the start of each task participants performed 10 practice trials to familiarise themselves with the task and instructions. For both the Acquire and Preserve tasks, the WM tasks were performed in the same experimental session as the reaching. However, in the case of the Acquire task the WM tasks were performed after the reaching task whereas for the Preserve task the WM tasks were performed first.

In the RWM (figure 3.1C, top row), the stimuli consisted of six 2D representations of 3D shapes drawn from an electronic library of the Shepard and Metzler type stimuli (Peters & Battista, 2008). The shape presented in the recall period was always the same 3D shape presented in the encoding period after undergoing a screen-plane rotation of 60°, 120°, 180°, 240° or 300°. In “match” trials, the only transform applied was the rotation; however, in “non-match” trials an additional vertical-axis mirroring was also applied. The difficulty of mental rotation has been demonstrated to increase with larger angles of rotation (Shepard & Metzler, 1971) and therefore the different degrees of rotation corresponded to the 5 levels of difficulty. However, given the symmetry of two pairs of rotations (60° and 300°, 120° and 240°), these 5 levels were collapsed to 3 for analysis.

In the SWM (figure 3.1C, middle row), stimuli in the encoding period consisted of a variable number of red circles placed within 16 squares arranged in a circular array (McNab & Klingberg, 2008). In the recall period, the array of squares was presented without the red circles and instead a question mark appeared in one of the squares. Participants then answered to the question “*Was there a red dot in the square marked by a question mark?*”

by pressing a corresponding key. In “match” trials the question mark appeared in one of the squares previously containing a red circle and in “non-match” trials it appeared in a square that was previously empty. Difficulty was scaled by varying the number of red circles (*i.e.* the number of locations to remember) from 3 to 7.

In the VWM (figure 3.1C, bottom row), participants were presented with a list of a variable number of consonants during the encoding period. In the recall period a single consonant was presented, and participants answered to the question “*Was this letter included in the previous array?*”. Thus, the letter could either be drawn from the previous list (“match” trials) or have been absent from the previous list (“non-match” trials). Difficulty in this task was determined by the length of the list to be remembered, ranging from 5 to 9.

Both the SWM and RWM tasks have been suggested to fall under the general umbrella term of spatial ability (Buszard & Masters, 2018). However, Miyake et al. (2001) suggest that although both mental rotation (*i.e.* RWM) and short term storage of spatial information (*i.e.* SWM) are within the spatial domain, RWM appears to rely more heavily on executive function and SWM on basic short term storage of spatial information. Furthermore, previous studies have found relationships between motor learning and this SWM task (Christou et al., 2016) and tasks similar to our RWM task (Anguera et al., 2010). Therefore, we included both tasks to assess if there was any severability in their relationships with reaching performance.

3.2.5 Genetic sample collection and profiling

COMT is thought to affect dopamine function mainly in the prefrontal cortex (PFC) (Egan et al., 2001; Goldberg et al., 2003), a region known for its involvement in WM and strategic planning (Anguera et al., 2010; Doll et al., 2015), whereas DARPP32 and DRD2 act mainly in the basal ganglia to promote exploitative behaviour, possibly by promoting selection of the action to be performed (Frank et al., 2009). Consequently, we focused here on single nucleotide polymorphisms (SNP) related to those genes: RS4680 (COMT) and RS907094 (DARPP32). Regarding DRD2, there are two potential SNPs available, RS6277 and RS1800497. Although previous studies focusing on exploration and exploitation have assessed RS6277 expression (Doll et al., 2016; Frank et al., 2009, 2007), it should be noted that this SNP varies greatly across ethnic groups, with some allelic variations being nearly completely absent in non-Caucasian-European groups (*e.g.* see RS6277 in 1000 Genomes Project (1000 Genomes Project Consortium et al., 2015)). This has likely been inconsequential in previous work, as Caucasian-European individual represented the majority of sampled groups; here however, this represents a critical shortcoming, as we aim at investigating a larger and more representative population including other ethnic groups. Consequently, we based our analysis on the RS1800497 allele of the DRD2 gene (Pearson-Fuhrhop et al., 2013).

At the end of the task, participants were asked to produce a saliva sample which was collected, stabilized and transported using Oragene DNA saliva collection kits (OG-500, DNAgenotek, Ontario, Canada). Participants were requested not to eat or drink anything except water for at least two hours before sample collection. Once data collection was completed across all participants, the saliva samples were sent to LGC (Hoddeson, Hertfordshire; `https://www.lgcgroup.com/`) for DNA extraction (per Oragene protocols: `https://www.dnagenotek.com/`) and genotyping. SNP genotyping was performed using the KASP SNP genotyping system. KASP is a competitive allele-specific PCR incorporating a FRET quencher cassette. Specifically, the SNP-specific KASP assay mix (containing two different, allele specific, competing forward primers) and the universal KASP master mix (containing FRET cassette plus Taq polymerase in an optimised buffer solution) were added to DNA samples and a thermal cycling reaction performed, followed by an end-point fluorescent read according to the manufacturer's protocol. All assays were tested on in-house validation DNA prior to being run on project samples. No-template controls were used, and 5% of the samples had duplicates included on each plate to enable the detection of contamination or non-specific amplification. All assays had over 90% call rates. Following completion of the PCR, all genotyping reaction plates were read on a BMG PHERAStar plate reader. The plates were recycled until a laboratory operator was satisfied that the PCR reaction had reached its endpoint. In-house Kraken software then automatically called the genotypes for each sample, with these results being confirmed independently by two laboratory operators. Furthermore, the duplicate saliva samples collected from 5% of participants were checked for consistency with the primary sample. No discrepancies between primary samples and duplicates were discovered.

3.2.6 Data analysis

Acquire task analysis

Reach trials containing movement times over 0.6s or less than 0.2s were removed from analysis (6.9% of trials). The end point angle of each movement was defined at the time when the radial distance of the cursor exceeded 10cm. This angle was defined in relation to the visible target with positive angles indicating clockwise rotations, end point angles and target angles for participants who experienced the counter clockwise rotations were sign-transformed. The explicit component of retention was defined as the difference between the mean reach angle of the maintain block and the remove block, while the implicit component was the difference between the mean reach angle of the remove block and baseline. If during the final 20 trials before the maintain block a participant achieved a mean reach angle within the reward region, participants were considered “successful” in learning the rotation; they were considered “unsuccessful” otherwise. For regression analysis a binary variable “task success” was defined as 1 and 0 for successful and unsuccessful participants, respectively. As in Holland et al. (2018), for unsuccessful participants, the largest angle of rotation at which the mean reach angle fell within the reward region was taken as the end of successful performance, and only trials prior to this point were included for further analysis. Success rate was defined as the percentage of trials during the learning blocks in which the end point angle was within the reward region. In order to examine the effect of reward on the change in end point angle on the subsequent trial, we calculated the absolute change in end point angle between consecutive trials (Holland et al., 2018; Sidarta et al., 2018; Therrien et al., 2016, 2018). Subsequently we calculated the median absolute change in angle following rewarded trials (ΔR) and the median absolute deviation of these values (MAD [ΔR]). This analysis was repeated for the changes in angle following unsuccessful trials (ΔP and MAD [ΔP]).

Preserve task analysis

Reach trials containing MTs over 1s were removed from analysis (2.38% of trials). The end point angle for each movement was defined at the time that the radial distance of the cursor from the start position exceeded 8cm. Trials in which the error was greater than 80° were excluded from further analysis (0.94% of trials). For each participant we plotted the reach error in one trial against the change in reach angle in the following trial for all trials in the adaptation block. The angle of the line of best fit was then used as the learning rate (Hutter & Taylor, 2018). Using this approach, a perfect adaptation leads to a value of -1, indicating that the error on a given trial is fully accounted for on the next trial. Overall this approach fitted the data well (mean $R^2 = 0.5038$, $SD = 0.12$). As in chapter 2, success rate, corresponding to percentage of rewarded trials, was measured separately in the first 30 and last 170 trials of the asymptote blocks and labelled early and late success rate, respectively. This reflects a dichotomy between a dominantly exploration-driven early phase and a later exploitation-driven phase. Implicit retention was defined as the difference between the average baseline reach direction and the mean reach direction of the last 20 trials of the last no-feedback block (Codol et al., 2018), similar to chapter 2. Analysis of changes in reach angle following rewarded trials were not pre-registered but were included *post-hoc*.

Exploratory analysis of reaching data

In order to understand which outcome variables in the reaching tasks were predictive of overall task success, we split the learning period into two sections for every participant. We assessed trial-by-trial changes in end point angle in the first section and compared them to success rate in the second section. For the Acquire task, we assessed trial-by-trial adjustments during the learning block, excluding the final 20 trials, and compared them to success rate in the last 20 trials of the learning block. In the Preserve task, we measured adjustments in the first 100 trials of the asymptote blocks and compared them to success rate in the last 100 trials of the asymptote blocks.

Two additional *post-hoc* stepwise regressions were performed on data from the Preserve task, including only data from participants with a success rate greater than 40% (N=70). Early and late success rate were the outcome variables and the same set of seven predictors as for the previous regressions were used (see section 3.2.7 for details).

Working memory tasks

WM performance was defined as the average percentage of correct responses across the 3 highest levels of difficulty for each task. In the case of the RWM task, the symmetrical arrangement of the angles of rotation in effect produced three levels of difficulty, and consequently all trials were analysed.

Genetics analysis

Genes were linearly encoded, with heterozygote alleles being 0, homozygote alleles bearing the highest dopaminergic state being 1, and homozygote alleles bearing the lowest dopaminergic state being -1 (table 3.1). All groups were assessed for violations of the Hardy-Weinberg equilibrium. The participant pool in the Preserve task was in Hardy-Weinberg equilibrium for all three genes considered, even when restricted to the Caucasian-only subpopulation. In the Acquire task population, COMT and DRD2 were in Hardy-Weinberg equilibrium, but DARPP32 was not ($p = 0.002$), with too few heterozygotes. Therefore, the DARPP32 alleles were recoded, with the heterozygotes (0) and the smallest homozygote group (C:C, -1) combined and recoded as 0. In the analysis including only the Caucasian subset, all three alleles were in the Hardy-Weinberg equilibrium, although combining the heterozygote and smallest homozygote group did not change the results.

SNP	Location	Allele coded as -1	Allele coded as 0	Allele coded as 1
rs4680	COMT	G:G (val:val) 31, 33	A:G (met:val) 68, 61	A:A (met:met) 17, 21
rsrs1800497	DRD2	T:T (lys:lys) 8, 7	T:C (lys:glu) 48, 51	C:C (glu:glu) 64, 62
rs907094	DARPP32	C:C 10, 21	C:T 54, 38	T:T 56, 62

Table 3.1: Coding for SNPs. The name of the SNP is provided along with the code assigned to each allele. The numbers represent the counts for the specific allele in the two tasks (Preserve, Acquire).

3.2.7 Statistical analysis

Regressions were performed using stepwise linear regressions (*stepwiselm* function in Matlab's *Statistics and Machine Learning Toolbox*), so as to select the most parsimonious model. In order to understand what genetic and WM factors are predictive of performance in the Acquire task, we performed a stepwise regression of the seven predictors (three allelic variations, three WM and ethnicity) onto each of several outcome measures representative of performance: success rate, implicit and explicit retention, ΔR, MAD[ΔR], ΔP, MAD[ΔP]. A stepwise logistic regression was employed for overall task success in the Acquire task. For the Preserve task, we performed separate stepwise regressions using the same seven predicators for the following outcome variables: baseline reach direction as a control variable, learning rate in the adaptation block, early and late success rate in the asymptote blocks (first 30 and last 170 trials; chapter 2; Codol et al., 2018), retention in the no-feedback blocks, and ΔR and ΔP during the asymptote blocks.

Prior to the regression analysis, all predictors and predicted variables were standardised (z-scored). For all non-ordinal variables, individual data were considered outliers if further than 3 standard deviations from the mean and were removed prior to standardisation. Multicollinearity of predictors was also assessed before regression with Belsley Collinearity Diagnositcs (*collintest* function in Matlab's *Econometrics Toolbox*) and no predictors were found to exhibit condition indexes over 30, indicating acceptable levels of collinearity. When considering retention for both tasks, unsuccessful participants were removed from the regression analysis.

In order to quantify the predictive ability of the regression models, a 10-fold cross-validation was performed on the model selected by the stepwise regression. Briefly, this consists of dividing the data samples into 10 evenly sized "folds". The data from nine of the folds are then used to create a regression model using ordinary least squares regression and this model is used to predict the values of data in the remaining fold given the values of the predictor variables. We measured the quality of the model fit in the 9 folds (in-sample) and the remaining fold of data (out-of-sample) by calculating the mean absolute error (MAE) of the predicted values from the real values. This process was repeated 1000 times for each model with the data assigned to each fold randomised on every iteration, we present the mean MAE ±SD across the 1000 iterations.

Mediation analysis methodology

We performed a mediation analysis to test if the relationship between SWM and success rate was mediated by ΔR. Our hypothesis was that higher SWM enables smaller changes after correct trials (ΔR), that is, to maintain performance more reliably, and that this explains the relationship between SWM and success rate. To ensure that separate trials were used in the calculation of ΔR and success rate, we split the trials into two equally sized folds. The success rate was then calculated for one fold as a percentage of correct trials, and ΔR was calculated as the median change of reach angle after correct trials in the other fold. For the Acquire task only successful subjects were included in the mediation analysis. We employed Baron and Kenny's three step mediation analysis (Baron & Kenny, 1986): first regress success rate on SWM, then regress ΔR on SWM, and finally regress success rate on both SWM and ΔR. Subsequently, we performed a Sobel test to determine if there was a significant reduction in the relationship between SWM and success rate when including ΔR. The Sobel test examines if the amount of variance in success rate explained by SWM is significantly reduced by including the mediator (Sobel, 1986). For a significant effect to be found, SWM must be a significant predictor of ΔR and ΔR must also be a significant predictor of success rate after controlling for the effect of SWM on success rate. We repeated this procedure 1000 times with the allocation of trials to each fold randomised on each repetition. We present results in terms of the 95% confidence intervals (CIs) for the R^2 values for each of the regressions and the median p-value of the Sobel test, along with the associated 95% CIs.

3.3 Results

3.3.1 Acquire task

In the Acquire task, participants had to learn to compensate for a secretly shifting reward region in order to obtain successful feedback (figures 3.2 and 3.3). As in Holland et al. (2018), about a quarter (28.1%) of participants failed to learn to compensate for the full extent of the rotation (figure 3.3A). Successful participants retained most of the learnt rotation (mean 80.7% ±28.0% SD) in the maintain block. However, upon being asked to remove any strategy they had been employing, their performance returned to near-baseline levels. Consequently, a large explicit component to retention was found for successful participants (figure 3.3B), whereas both successful and unsuccessful participants manifest a small but non-zero implicit component ($t(86) = 9.90, p = 7.43e^{-16}, d = 1.061$ and $t(33) = 4.53, p = 7.39e^{-5}, d = 0.776$, respectively; figure 3.3C). Furthermore, in accordance with Holland et al. (2018), we found that participants made larger ($t(120) = 15.80, p = 4.32e^{-31}, d = 1.900$) and more variable changes in reach angle following unrewarded trials compared to rewarded trials ($t(120) = 13.36, p = 1.68e^{-25}, d = 1.485$; figure 3.3D-G). Comparing participants who would go on to fail to those who will not, the post-error adjustments were smaller for failing participants than in successful participants (independent t-test: $t(119) = 3.33, p = 0.001, d = 0.672$; Figure 3D). However, changes following rewarded trials were similar between the groups (independent t-test: $t(119) = 0.71, p = 0.48, d = 0.143$; figure 3.3F, G). The results obtained with this sample size (N=121) therefore replicate results from a previous study (N=30) and provides further confirmation that performance in this task is fundamentally explicitly driven (Holland et al., 2018).

In order to understand what genetic and WM factors are predictive of performance in the reaching task, we performed a stepwise regression of the seven predictors (three allelic variations, three WM and ethnicity) onto each of several outcome measures representative of performance: success rate, implicit and explicit retention, ΔR, MAD[ΔR], ΔP, MAD[ΔP]. Additionally, we performed a stepwise logistic regression of the predictors onto a binary variable encoding if a participant successfully learnt the full rotation (1) or not (0), and that we labelled *task success.* The logistic regression showed no significant predictors of task success, that is, of being able to follow the shifting reward region until the end of the learning block. However, higher SWM was predictive of an increased success rate (percentage of correct trials; $\beta = 0.416, p = 2.95e^{-6}$). To ensure that the relationship between SWM and success rate was not due to failure at a later point in the task, success rate was measured during the initial period in which subjects who could not fully account for the displacement are still successful; for those who could, the full

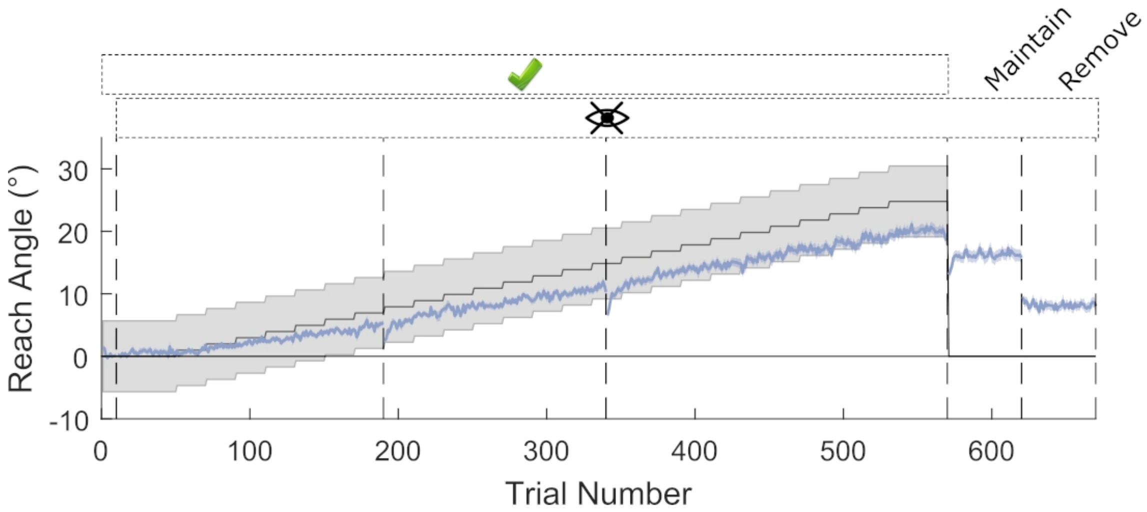


Figure 3.2: Reaching performance in the Acquire task. The grey region represents the gradually rotating rewarded region, the blue line represents mean reach angle for each trial, and the shaded blue region represent s.e.m. Vertical dashed lines represent different experiment blocks or breaks. The rectangle enclosing the green tick above the axes represents trials in which reward was available, and the rectangle with the "eye" symbol indicates when vision was not available. Average performance for the full cohort falls within the reward region and demonstrates a clear reduction in reach angle when asked to remove strategy. N=121.

learning block was included. Next, retention was assessed by splitting up the explicit and implicit components such as in Holland et al. (2018). No predictor was related to the implicit component, but the explicit component was strongly and positively associated with RWM ($\beta = 0.373, p = 1.78e^{-4}$). These results suggest positive relationships for both RWM and SWM with task performance: greater RWM predicts a greater contribution of explicit processes to learning, whereas greater SWM predicts a greater percentage of correct trials.

In Holland et al. (2018), the amplitude of the changes in reach angle participants made following unrewarded trials was found to be predictive of task success, that is, greater ΔP was predictive of an increased chance of overall task success. Thus it could be that RWM and SWM, that are shown to associate with performance in this study, are themselves predictors of changes in reach angle. The regression results demonstrated that a large ΔR was inversely related to SWM ($\beta = -0.251, p = 0.006$), as was MAD[ΔR] ($\beta = -0.283, p = 0.002$). The results indicate that greater SWM was predictive of smal-

Population	Outcome	Predictor	β±SE	p	Model	MAE	
						In-sample	Out-of-sample
All	SR	SWM	0.416 ± 0.085	$2.954e^{-6}$	$F(113, 2) = 12.280, p = 1.493e^{-5}$	$0.712 \pm 3.627e^{-4}$	0.734 ± 0.007
		COMT	0.087 ± 0.084	0.303			
	Explicit	RWM	0.373 ± 0.095	$1.784e^{-4}$	$F(85, 2) = 15.370, p = 1.78e^{-4}$	$0.789 \pm 6.553e^{-4}$	0.810 ± 0.012
	MAD(ΔP)	RWM	-0.236 ± 0.091	0.011	$F(116, 2) = 6.767, p = 0.011$	$0.713 \pm 3.202e^{-4}$	0.727 ± 0.006
	ΔR	SWM	-0.251 ± 0.090	0.006	$F(116, 2) = 4.420, p = 0.014$	$0.771 \pm 4.380e^{-4}$	0.787 ± 0.006
	MAD(ΔR)	SWM	-0.283 ± 0.090	0.002	$F(116, 2) = 5.292, p = 0.006$	$0.749 \pm 2.986e^{-4}$	0.763 ± 0.006
Caucasian	SR	SWM	0.283 ± 0.101	0.006	$F(80, 2) = 7.882, p = 0.006$	$0.761 \pm 5.629e^{-4}$	0.782 ± 0.008
	Explicit	RWM	0.300 ± 0.105	0.006	$F(53, 2) = 8.207, p = 0.0064$	$0.741 \pm 9.733e^{-4}$	0.773 ± 0.019
	MAD(ΔP)	RWM	-0.237 ± 0.109	0.033	$F(78, 2) = 4.730, p = 0.033$	$0.759 \pm 4.767e^{-4}$	0.779 ± 0.009
	ΔR	SWM	-0.207 ± 0.101	0.044	$F(78, 2) = 4.188, p = 0.044$	$0.751 \pm 5.852e^{-4}$	0.775 ± 0.009
	MAD(ΔR)	SWM	-0.215 ± 0.105	0.044	$F(78, 2) = 4.176, p = 0.044$	$0.757 \pm 4.098e^{-4}$	0.777 ± 0.009

Table 3.2: Regressions with significant models for the Acquire task. The predictors selected by the stepwise regression procedure to have a model significantly better than the intercept only model are reported. For each model the selected predictors are reported alongside the coefficient and standard error and associated p value for that predictor, as well as the significance of the model overall. The results of the 10-fold cross-validation analysis are reported in terms of the mean ±SD of the absolute error (MAE) of the model prediction for the 1000 repetitions. Results are reported when including all participants (N=121) or the Caucasian only subset (N=82), demonstrating that the reported results are consistent in both. SR: success rate.

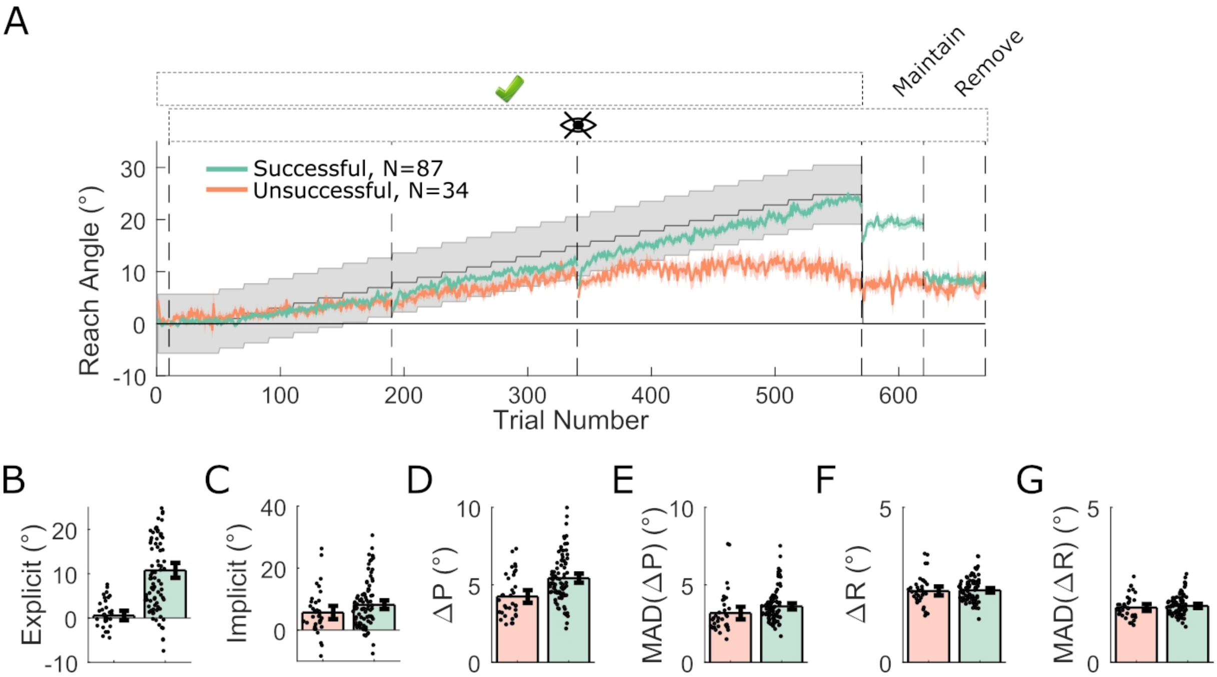


Figure 3.3: Acquire task split by success at final angle. A. Average reach angle for the successful (green) and unsuccessful (orange) groups, shaded regions represent s.e.m. and grey shaded region represents the rewarded region. The rectangle enclosing the green tick above the axes represents trials in which reward was available, and the rectangle with the "eye" symbol indicates when vision was not available. Despite similar initial performance, a clear divergence can be seen between the two groups and an explicit component to retention is only visible in the successful group, whereas implicit retention is similar between groups. B-G. subplots displaying derived measures, which acted as outcome variables for the regression analysis, separated into successful and unsuccessful participants overlaid with individual data points. Error bars represent 95% bootstrapped CIs.

ler and less variable changes in reach angle after successful trials, suggesting high SWM enables the maintenance of rewarding reach angles. It was also found that the variability of changes in reach angle after unrewarded trials (MAD[ΔP]) was negatively predicted by RWM ($\beta = -0.236, p = 0.011$). This result was unexpected as it suggests that greater WM capacity predicts smaller changes following unrewarded trials, whereas previous results suggest a positive relationship between these changes and overall task success. Finally, to ensure the robustness of the results, we tested whether retaining only the largest ethnic group in our population (*i.e.* Caucasian, N=82/121) produced the same results as was observed with the full participant pool. In accordance with the full sample, all previously described relationships were also found in the Caucasian only subset (table 3.2).

Overall, WM (in particular RWM and SWM) successfully predicted various aspects of performance in the Acquire task, while genetic predictors failed to do so. Specifically, greater SWM predicted smaller and less variable changes after correct trials. This suggests that SWM underlies one's capacity to preserve and consistently express an acquired reach direction to obtain reward. Furthermore, SWM also directly predicted success rate. In addition, greater RWM was a strong predictor of explicit control. The inverse relationship between RWM and the variability of changes after unrewarded trials was unexpected. However, one possible explanation is that participants with poorer WM capacity make larger errors which require larger corrections. Restricting our group to Caucasians showed that these effects are robust to ethnicity.

3.3.2 Preserve task

In this task, we addressed how well participants can maintain a previously learnt adaptation after transitioning to binary feedback. As participants are unable to compensate for a large abrupt displacement of a hidden reward region (van der Kooij & Overvliet, 2016; Manley et al., 2014), participants first adapted to an abruptly introduced 20° clockwise rotation with full vision of the cursor available. Subsequently, visual feedback of the cursor position was replaced with binary feedback; participants were rewarded if they continued reaching towards the same angle that resulted in the cursor hitting the target during the adaptation phase. Overall, participants adapted to the visuomotor rotation successfully (figure 3.4 and 3.5A-C) before transitioning to the binary-feedback asymptote blocks. However, from the start of the asymptote blocks onward, participants exhibited very poor performance, expressing an average 45.0% ±24.2 SD success rate when considering all 200 asymptote trials (figures 3.4 and 3.5A, D, E). Separating successful and unsuccessful participants (40% success rate cut-off; figure 3.5A) revealed that successful participants expressed behaviour greatly similar to that observed in chapter 2, in which unsuccessful participants were excluded, using the same cut-off (40% success rate). The

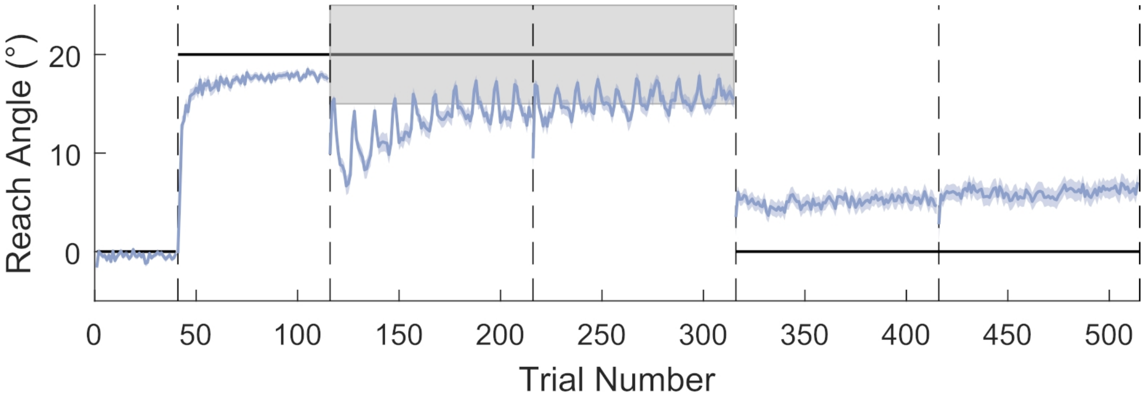


Figure 3.4: Reaching performance in the Preserve task. The grey shaded area represents the rewarded region, and the thick black line represents the perturbation. The vertical dashed lines represent block limits. The blue line indicates mean reach angle for every trial and blue shaded areas represent s.e.m. After successfully adapting to the visuomotor rotation performance deteriorates at the onset of binary feedback, subsequently success rate increases towards the end of the asymptote blocks. Following the removal of all feedback, and the instruction to remove any strategy, a small amount of implicit retention remains. N=120.

"spiking" behaviour observed in reach angles during the asymptote blocks (figure 3.5A) is due to the presence of the "refresher" trials, with the large positive changes in reach angle corresponding to trials immediately following the refresher trials. This pattern of behaviour is particularly pronounced in the unsuccessful participants. Finally, participants demonstrated at least a residual level of retention even though they were being instructed to remove any strategy they had employed ($t(69) = 7.268, p = 3.345e^{-10}, d = 0.869$; figure 3.5A, F). Therefore, the results obtained in this sample (N=120) replicate results from chapter 2 (Codol et al., 2018; N=20, BF-Remove group) and provides further confirmation that performance in this task is fundamentally explicitly driven. As with the Acquire task, successful participants displayed larger changes in angle after unrewarded trials than their unsuccessful counterparts ($t(117) = 3.847, p = 1.952e^{-4}, d = 0.717$; figure 3.5H). However, in contrast to the Acquire task, successful participants also displayed smaller changes in reach angle after rewarded trials ($t(115) = -7.534, p = 1.218e^{-11}, d = 1.421$; figure 3.5G).

As in the Acquire task, we examined if performance in any of the WM tasks or genetic profile could predict participant's behaviour in the reaching task. We performed separate stepwise regressions for the following outcome variables: baseline reach direction as a

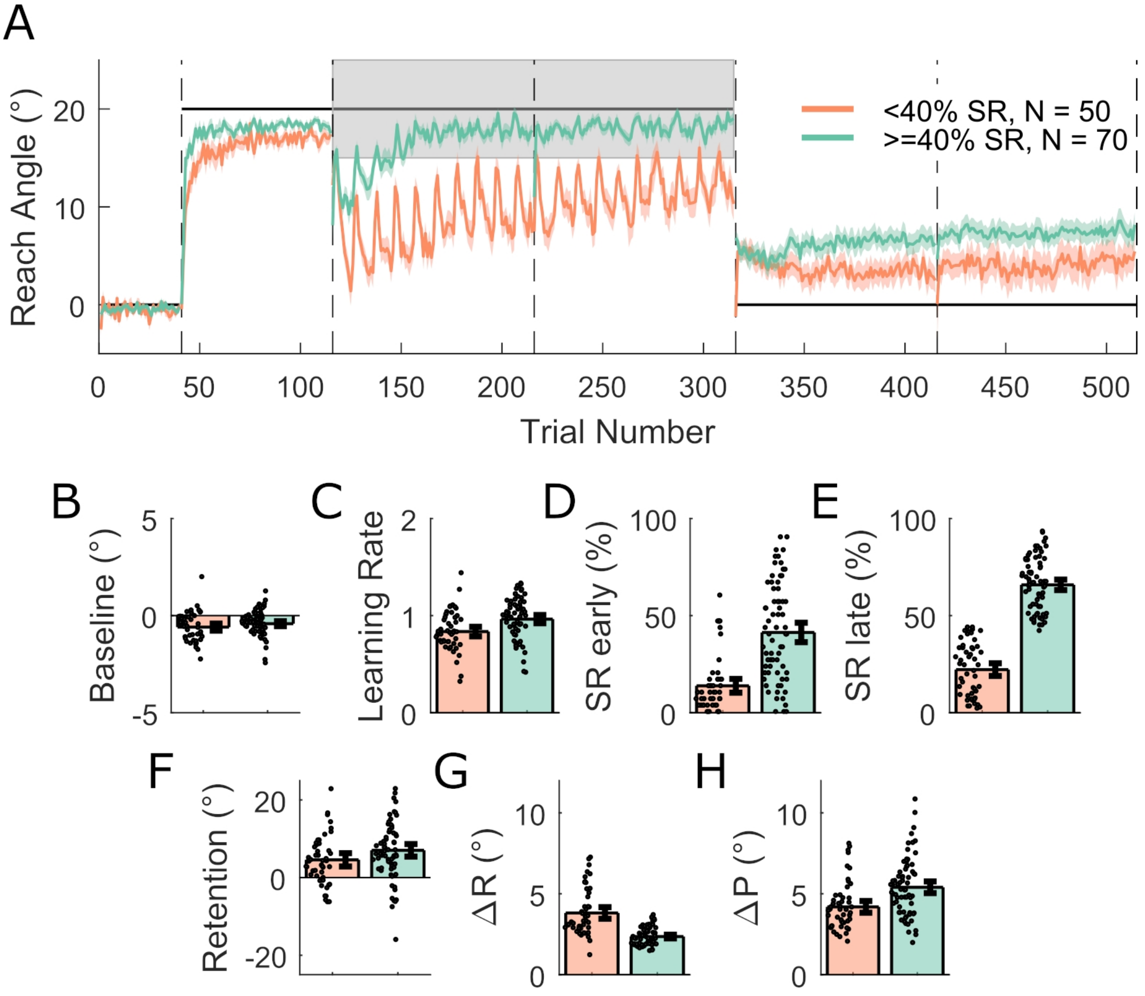


Figure 3.5: Preserve task split into two groups on the basis of success rate. A. Shaded regions represent s.e.m. B-H. Derived variables, which acted as outcome variables for the regression analysis, for the two groups, error bars on the bars represent 95% bootstrapped CIs and individual data points are displayed. SR: success rate.

control variable, learning rate in the adaptation block, early and late success rate in the asymptote blocks (first 30 and last 170 trials, similar to chapter 2), retention in the no-feedback blocks, and ΔR and ΔP during the asymptote blocks. The most striking result was that both early and late success rate could be reliably predicted by RWM (early: $\beta = 0.255, p = 0.005$; late: $\beta = 0.287, p = 0.002$; table 3.3), with greater RWM associated with increased success rate.

Genetic profile did not predict any aspect of performance, analogous to the Acquire task. In contrast, greater SWM successfully predicted reduced ΔR ($\beta = -0.194, p = 0.036$), similarly to the Acquire task. Finally, retention values were surprisingly predicted by eth-

						MAE	
Population	Outcome	Predictor	β±SE	p	Model	In-sample	Out-of-sample
All	early SR	RWM	0.255 ± 0.089	0.005	$F(2,118) = 8.207, p = 0.005$	$0.814 \pm 3.416e^{-4}$	0.830 ± 0.005
	late SR	RWM	0.287 ± 0.088	0.002	$F(2,118) = 10.583, p = 0.002$	$0.800 \pm 3.984e^{-4}$	0.816 ± 0.005
	Retention	Ethnicity	-0.528 ± 0.248	0.037	$F(2,68) = 4.525, p = 0.037$	$0.715 \pm 7.025e^{-4}$	0.741 ± 0.022
	ΔR	SWM	-0.194 ± 0.091	0.036	$F(2,118) = 4.502, p = 0.036$	$0.707 \pm 4.178e^{-4}$	0.721 ± 0.006
Caucasian	late SR	RWM	0.232 ± 0.106	0.031	$F(2,83) = 6.766, p = 0.011$	$0.804 \pm 3.884e^{-4}$	0.827 ± 0.008
	Retention	DARPP32	-0.214 ± 0.101	0.040	$F(2,45) = 4.451, p = 0.040$	$0.529 \pm 6.248e^{-4}$	0.554 ± 0.027
SR >40%	late SR	SWM	0.156 ± 0.069	0.026	$F(2,68) = 5.173, p = 0.026$	$0.434 \pm 4.720e^{-4}$	0.449 ± 0.005

Table 3.3: Regression with significant models for Preserve task. The predictors selected by the stepwise regression procedure to have a model significantly better than the intercept only model are reported. For each model the selected predictors are reported alongside the coefficient and standard error and associated p value for that predictor, as well as the significance of the model overall. The results of the 10-fold cross-validation analysis are reported in terms of the mean ±SD of the absolute error (MAE) of the model prediction for the 1000 repetitions. Results are reported when including all participants (N=120) or the Caucasian only subset (N=85), demonstrating that the relationship between RWM and late success rate are consistent in both and revealing a genetic predictor of retention. SR: success rate.

nicity ($\beta = -0.528, p = 0.037$). Due to the existence of a relationship between ethnicity and retention, we performed the same analysis as in the Acquire task, that is, we tested if our observed results hold if only our largest ethnic group (Caucasian, N=85/120) was considered. In accordance with the result for the full population, the positive relationship between late success rate and RWM was again observed ($\beta = 0.232, p = 0.031$). However, the SWM-ΔR and RWM-early success rate relationships were no longer observed in this smaller subset of the population. Interestingly, retention was now predicted by a genetic variable, DARPP32 ($\beta = -0.214, p = 0.040$), suggesting that less dopaminergic metabolism leads to better retention. This last result again suggests a possible confound, that is, that DARPP32 allelic distribution was different across ethnic groups. However, a χ^2 test analysis demonstrated that DARPP32 alleles were evenly distributed between the Caucasian and non-Caucasian group, ruling out this possibility ($\chi^2 = 2.578, p = 0.276$). As a *post-hoc* analysis we performed the same stepwise regressions for the outcome variables early and late success rate but restricted to participants with an overall success rate of greater than 40%. Although we found no predictors of early success rate, we did find that higher SWM was predictive of a higher late success rate ($\beta = 0.156, p = 0.026$). This result is in contrast to the relationship of RWM to late success rate when including all participants.

Overall the regression results fit a pattern similar to that found for the Acquire task with greater RWM predicting improved performance on the reaching task and greater SWM predicting smaller changes in reach angle after rewarded trials. However, we observed a genetic predictor of performance in one specific instance in the Preserve task.

3.3.3 Cross-validation analysis

To test the predictive ability of the regression models we performed 10-fold cross-validation on the final model selected by the stepwise regression process. The quality of the in-sample and out-of-sample fits was assessed by calculating the MAE. From tables 3.2 and 3.3, it can be seen that although the out-of-sample MAE is consistently higher than that of the corresponding in-sample, all differences are less than 0.1 and all of out-of-sample MAEs are below 1. As both the predictor and outcome variables are standardised this indicates that the mean error of the predicted outcome value was less than 1 standard deviation away from the true outcome value, and the small magnitude of increases observed between the in-sample and out-of-sample indicates that the models make accurate predictions when confronted with data on which they were not trained.

3.3.4 Exploratory analysis

As a relationship exists between SWM and ΔR in both the Acquire and Preserve paradigms, we ran exploratory regressions to assess the relationship between ΔR and success rate across both tasks. Since ΔR and success rate are conceptually strongly related variables, and measuring on the same dataset would render them non-independent, we split each individual's reaching data into two sections and assessed whether ΔR or ΔP in the first section could reliably predict success rate in the second (see methods section 3.2.6 for details). Although we had not found no predictors of ΔP in our primary analysis, results here as well as in previous work (Holland et al., 2018) has demonstrated a link between ΔP and task success, with a greater ΔP indicative of greater success. Therefore, we also performed the same analysis for ΔP.

In the Acquire task, ΔR and ΔP in the first section of learning trials predicted success rate in the final twenty trials, though ΔP appeared as the strongest predictor (ΔR: $\beta = -0.274, p = 0.015$; ΔP: $\beta = 0.581, p = 3.89e^{-6}$; figure 3.6A, B, yellow; table 3.4). Similarly, for the Preserve task ΔR and ΔP in the first half of asymptote trials predicted success rate in the second half (ΔR: $\beta = -0.750, p = 1.07e^{-12}$; ΔP: $\beta = 0.229, p = 0.007$; figure 3.6A, B, pink; table 3.4). In both tasks, the directions of these relationships were opposite; greater success rate was predicted by smaller changes following rewarded trials and greater changes following unrewarded trials. In summary, we found that for both

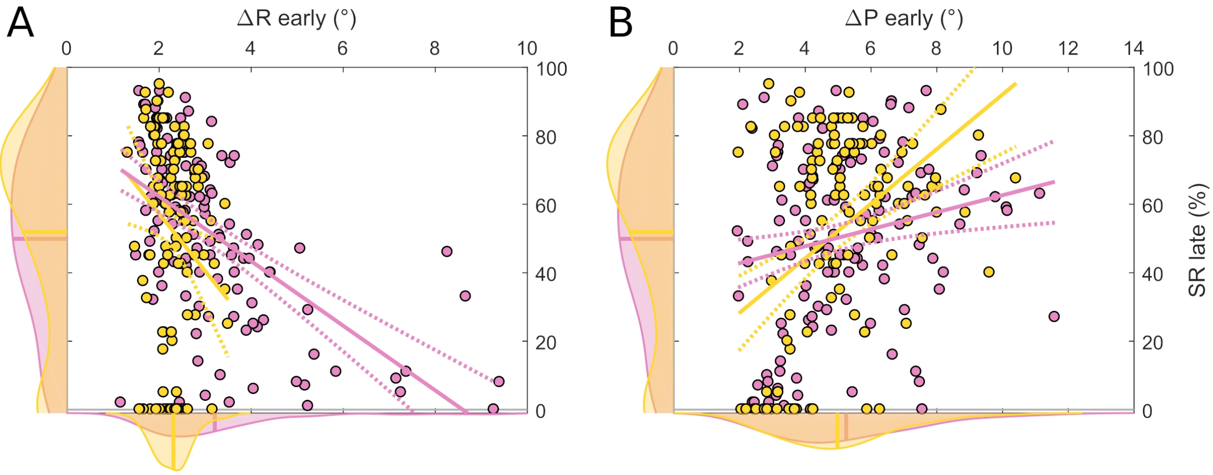


Figure 3.6: Slice plots of regression results for prediction of late success rate by changes in reach angle. Panels indicate results following rewarded (A) and unrewarded (B) trials during the early learning period. The central axis of each panel displays the individual data from the Acquire (yellow) and Preserve (pink) task, the smoothed distributions of the data in each dimension are displayed on the corresponding axes. Solid lines represent the prediction of the regression model when the other predictor is held at its mean value, and dashed lines represent CIs. SR: success rate.

tasks the magnitude of changes in behaviour in response to rewarded and unrewarded trials early in learning were strongly predictive of future task success across both the Acquire and Preserve tasks.

		ΔR	ΔP	Model
Acquire	β	-0.274	0.581	$F(115,2) = 11.9, p = 2.09e^{-5}$
	SE	0.111	0.120	
	p	0.015	$3.89e^{-6}$	
Preserve	β	-0.750	0.229	$F(112,2) = 35.3, p = 1.28e^{-12}$
	SE	0.093	0.084	
	p	$1.07e^{-12}$	0.007	

Table 3.4: Regression results for split data for both the Acquire and Preserve tasks. Ordinary least squares linear regressions were performed with both ΔR and ΔP included as predictors. The regression coefficient, standard error and p value for each predictor are reported along with the significance of the comparison between the model and an intercept only model. In both tasks there is an opposing relationship between ΔR and ΔP and SR, with smaller changes after rewarded trials and larger changes after unrewarded trials predictive of success.

3.3.5 Mediation analysis

Finally, to test whether the effect observed between SWM and success rate was explained by an indirect effect through ΔR, we performed an exploratory mediation analysis on both tasks (figure 3.7). For both the Acquire and Preserve tasks, the results indicate a significant proportion (R^2 = 6.13–22.14% of total variance, median $p = 7.10e^{-4}$ and R^2 = 1.24–4.51% of total variance, $p = 0.04$, respectively) of the effect of SWM on success rate can be explained by a mediation from SWM via ΔR to success rate. However, in the case of the Acquire task, a significant relationship of SWM on success rate also remained, indicating that not all of the effect of SWM on success rate could be explained by the indirect pathway. Of note, in the Preserve task the indirect mediation SWM-to-ΔR was weaker and was not significant on every repetition, occasionally leading to an insignificant mediation effect despite the median p-value indicating an effect when considering all repetitions.

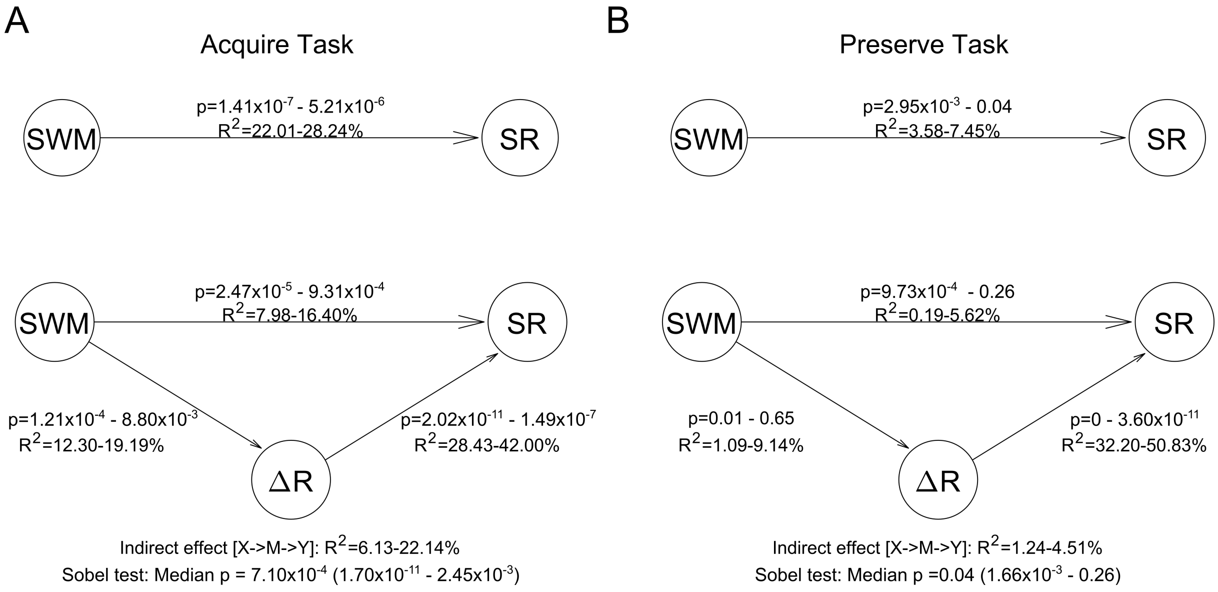


Figure 3.7: Mediation Analysis for both the Acquire and Preserve tasks. A. Acquire task. B. Preserve task. The numbers associated with each arrow display the 95% CIs for each of the relationships (R^2 and p-values) across the 1000 repetitions. Below the figure, the results of the Sobel test are displayed indicating the amount of variance explained by the indirect pathway and the 95% CIs and median p-value. SR: success rate.

3.4 Discussion

In this study, we sought to identify if genetic background or specific domains of WMC could explain the variability observed in performance levels during reward-based motor learning tasks. We found that RWM and SWM predicted different aspects of the Acquire and Preserve tasks, whereas VWM did not relate to any performance measure. Specifically, RWM predicted the explicit component of retention in the Acquire task and success rate in the Preserve task, whereas SWM predicted success rate in the Acquire task and ΔR in both tasks. Conversely, allelic variations of the three dopamine-related genes (DRD2, COMT and DARPP32) did not consistently predict any behavioural variables in the full sample of participants. This suggests that SWM predicts a participant's capacity to reproduce a rewarded motor action, while RWM predicts a participant's ability to express an explicit strategy when making large behavioural adjustments. Therefore, we conclude that WMC plays a pivotal role in determining individual ability in reward-based motor learning.

3.4.1 Spatial and mental rotation working memory have a dissociable but partially overlapping role in reward-based motor learning

Recently, Wong et al. (2019) described a positive relationship between SWM and the development of explicit strategies in visuomotor adaptation, complementing previous reports (Anguera et al., 2012; Christou et al., 2016). However, in contrast to the current findings, the previous experiments employed relatively small sample sizes, which may render correlations unreliable. The large group sizes employed here and the confirmation of relationships across two tasks provides strong evidence that these relationships are robust, replicable, and extend from visuomotor adaptation to reward-based motor learning. An interesting dichotomy was the reliance on SWM and RWM for the Acquire and Preserve task, respectively. While the Preserve task required the maintenance of a large, abrupt behavioural change, the Acquire task required the gradual adjustment of behaviour considering the outcomes of recent trials. Therefore, RWM may underscore one's capacity to express a large correction consistently over trials with binary feedback, whereas SWM reflects one's capacity to maintain a memory of previously rewarded actions and adjust behaviour accordingly. Conformingly, the magnitude of ΔR was negatively related to SWM but not RWM in both tasks, suggesting high SWM enables the maintenance of rewarding actions. Supporting this, Sidarta et al. (2018) reported that higher SWM was associated with reduced movement variability in a reward-based motor learning task. Additionally, explicit retention, an element of the Acquire task that requires a large, sudden change in reach direction, was predicted by RWM rather than SWM. Finally, it should be noted that RWM and SWM were often selected as predictors simultaneously. The overlapping but distinct pattern of relationships between RWM, SWM and outcome measures considered here supports the view that they share substrates at least partially, but have different patterns of dependency on executive functions (Miyake et al., 2001), explaining why differences can be observed as well.

3.4.2 Behavioural performance in the Preserve underline a lack of generalisation from “refresher” trials

A notable feature of the Preserve task is the “spiking” behaviour observed immediately following “refresher” trials, suggesting a central role of refresher trials in binary feedback-based performance when included (Codol et al., 2018; Shmuelof et al., 2012). The transient nature of this decrease in error demonstrates these trials are insufficient to promote generalisation to binary feedback trials, at least in unsuccessful participants. It remains an open question whether superior performance of successful participants was partly due to a capacity to generalise information from “refresher” trials. McDougle & Taylor (2019) provided evidence that two separate strategies are employed in visuomotor adaptation: response-caching and mental rotation. The balance between the two strategies is a function of task demands. It could be that the relationships between RWM and SWM to success rate in the Preserve and Acquire tasks, respectively, reflect a different balance of the use of these strategies. Visual feedback in refresher trials in the Preserve task would encourage the engagement of mental rotation processes, whereas the slow updating of behaviour in the Acquire task engages the response-caching memory system. Interestingly, this would imply that response-caching is associated with SWM.

3.4.3 Reliance on working memory suggests the use of explicit control for reward-based motor learning

Surprisingly, although ΔP was a strong predictor of success in both tasks, it was not predicted by any genetic variable. In the Acquire task, MAD[ΔP] was inversely predicted by RWM. This result is surprising given the positive relationship between ΔP and success rate in both tasks. Furthermore, although no predictor of ΔP was found in the Preserve task, ΔP should be important for explicit control, as errors are a central element leading to the induction of structural learning in reward-based tasks, reinforcement learning (Daw et al., 2011; Manley et al., 2014; Sutton & Barto, 1998) and motor learning in general (Maxwell et al., 2001; Sidarta et al., 2018). In line, in the Acquire task, we observe that breaks have a strong influence on performance, leading to a drop in performance that successful—but not unsuccessful participants—quickly recover from (figure 3.3A). IT may be possible that those breaks speed up the process of dissociating successful and unsuccessful participants. Similarly, in the Preserve task, the break between the first and second assymptote block shows that successful participants have to recover from a small drop in performance as well. However, since unsuccessful participants have already failed at that point, it may not promote a dissociation between the two pools of participants as in the Acquire task. These possibilities may be tested in the future by altering the position and number of breaks.

If RWM is important for explicit control and the main element predicting success in the Preserve task, it may be worth considering whether gradual designs (as in the Acquire task) are more suitable to engage implicit reinforcement learning, at least initially. However, the Acquire task still bears a strong explicit component (Holland et al., 2018). How can these two views be reconciled? In reward-based motor learning, it is observed that participants begin to reflect on task structure and develop strategies upon encountering negative outcomes (Leow et al., 2016; Loonis et al., 2017; Maxwell et al., 2001), which occurs nearly immediately in the Preserve task after the introduction of binary feedback, due to a lack of generalisation of cerebellar memory (Codol et al., 2018). In contrast, in the Acquire task, participants experience an early learning phase with mainly rewarding outcomes, possibly suppressing development of explicit control and allowing for this early window of implicit reward-based learning. It may be possible to assess the effect of reinforcement, or specifically rewarding and punishing outcomes on motor learning in futures studies using a closed-loop design, whereby the amount of reward or punishment is a direct function of performance (Therrien et al., 2016). Other studies have demonstrated that minor adjustments in reach direction under reward-based feedback can occur, though none has assessed their explicitness directly in the very early stages, such as about

1-4° (Izawa & Shadmehr, 2011; Pekny et al., 2015; Therrien et al., 2016). Notably, Izawa & Shadmehr (2011) observed that after 8° shifts of a similarly-sized reward region, participants indeed noticed the perturbation, but awareness was not assessed for earlier, smaller shifts. Future studies may be able to assess explicit control during small shifts by assessing proprioceptive bias or using landmark reporting on a trial-by-trial basis. Alternatively, preventing the use of explicit control by using a dual task may allow observing the exclusive contribution of implicit learning (Holland et al., 2018; Haith et al., 2015). Finally, the consistent remaining implicit component may reflect a combination of implicit reinforcement (Shmuelof et al., 2012), use-dependent plasticity (Bütefisch et al., 2000; Classen et al., 1998), perceptual bias (Vindras et al., 1998) or even perceptual recalibration (Modchalingam et al., 2019) and is similar to that observed in the feedbacl-instruction experiment in chapter 2. This may suggest that the time spent on the task—*i.e.* the amount of trials—will alter this effect, a possibility that can be easily tested by running a similar experiment while manipulating the amount of trials in the learning block.

In Holland et al. (2018), the addition of a RWM-like dual-task proved very effective in preventing explicit control, leading to participants invariably failing at the reaching task. Therefore, it may seem surprising that RWM does not predict success rate in the Acquire task. A possible explanation is that RWM and SWM share the same memory buffer, as was mentioned earlier (Anguera et al., 2010; Beschin et al., 2005; M. S. Cohen et al., 1996; K. Jordan et al., 2001; Suchan et al., 2006; Miyake et al., 2001). Similarly, in force-field adaptation the early component of adaptation—considered as bearing a strong explicit element—is selectively disrupted with a VWM dual-task (Keisler & Shadmehr, 2010). However, we found no relationships with VWM in our reward-based motor tasks. It may be possible that reward-based motor performance relies more on spatial instances of WM as opposed to tasks such as force-field adaptation.

3.4.4 Implications of the null effect for genetic predictors

The absence of dopamine-related genetic relationships with behaviour is a surprising result as a substantial body of literature points to a relationship between dopamine and performance in reward-based tasks, including those with motor components (Deserno et al., 2015; Doll et al., 2016; Frank et al., 2009, 2007; Gershman & Schoenbaum, 2017; Izawa & Shadmehr, 2011; Nakahara & Hikosaka, 2012; Pekny et al., 2015; Therrien et al., 2016), and there is a growing appreciation of the links between decision-making and motor learning (Chen, Holland & Galea, 2018; Haith & Krakauer, 2013). For instance, Chen et al. (2017b) demonstrated that exploratory motor learning can be modelled as a sequential decision-making process, with explorative drive being shared between motor and decision-making tasks. However, the results presented here suggest that genetic predictors of exploration and exploitation in decision-making tasks are not also predictive of similar behaviours in reward-based motor learning.

A possibility is that our study missed an existing effect due to lack of statistical power. However, our sample sizes were defined *a priori* for 90% power based on previous work (Doll et al., 2016; Frank et al., 2009; see pre-registrations online as detailed in methods, section 3.2), and are therefore unlikely to be underpowered. Another possibility is that we employed the wrong variables to assess behaviour. However, given the informative and coherent relationships between WM and motor learning, it could be that the allelic variability of the genes we selected does not impact performance in reward-based tasks in any meaningful way, either because their downstream consequences are negligeable or because they are easily compensated by other mechanisms such as WM itself. In line with this, a recent study showed that dopamine pharmacological manipulation did not alter reward effects in a visuomotor adaptation task (Quattrocchi et al., 2018). However, previous work has shown that Parkinson's disease patients show impaired reward-based motor performance (Pekny et al., 2015). This is in line with the supposition that genetic variations may not impact reward-based motor learning significantly by themselves, while the wide depletion of dopaminergic neurons in Parkinson's disease would. Finally, future work should also address the possible involvement of other neuromodulators, such as acetylcholine, norepinephrine and serotonin (for a review, see Dash et al., 2007), in reward-based motor learning.

3.4.5 Conclusions

In summary, despite employing two distinct tasks and an independent participant pool on different devices, we find strikingly similar results across both tasks regarding reward-based motor learning. While SWM strongly predicted a participant's capacity to reproduce successful motor actions, RWM predicted a participant's ability to express an explicit strategy when required to make large behavioural adjustments. Therefore, both SWM and RWM are reliable predictors of success during reward-based motor learning. Surprisingly, no dopamine-related genotypes predicted performance. Therefore, WMC plays a pivotal role in determining individual ability during reward-based motor learning. This could have important implications when using reward-based feedback in applied settings, as our study suggests that only a subset of the population may benefit from such approach.

Chapter 4

REWARD-BASED IMPROVEMENTS IN MOTOR CONTROL ARE DRIVEN BY MULTIPLE ERROR-REDUCING MECHANISMS

In this chapter, the hypothesis for the reward-magnitude experiment, reward-punish experiment and start-of-reach stiffness experiment were defined and the paradigms were designed conjointly by P.J. Holland, J.M. Galea and me. The hypothesis for the end-of reach stiffness experiment was defined and the paradigm designed by me. All experimental dataset were acquired and analysed by me. The time-time correlation analysis was suggested by S.G. Manohar and done by me. The simulations were designed, implemented and quantitatively tested by me, with additional help from S.G. Manohar. All interpretations of results were done conjointly by P.J. Holland, J.M. Galea, S.G. Manohar and me. J.M. Galea provided the fundings and materials. I drew the figures and wrote this chapter.

4.1 Introduction

Motor control involves two main components that can be both optimised; action selection and action execution (Chen, Holland & Galea, 2018). While the former addresses the problem of finding the best action to achieve a goal amongst a subset of actions, the latter is concerned with performing the selected action with the greatest precision possible (Chen, Holland & Galea, 2018; Shmuelof et al., 2014; Stanley & Krakauer, 2013). Naturally, both processes come at a computational cost, meaning the faster an action is selected or executed, the more prone it is to errors—a phenomenon formalised as Fitts' law (Fitts, 1954). This is represented in a speed-accuracy function where accuracy decays as speed increases. Because speed-accuracy functions are a hallmark of human limitation in motor control, they have been regularly used to quantify performance (Reis et al., 2009; Telgen et al., 2014). For example, in skill learning, one may see the speed-accuracy function shift so that higher levels of accuracy are observed for any given speed (Reis et al., 2009; Telgen et al., 2014).

Interestingly, both action selection and action execution are highly susceptible to the presence of reward. For instance, introducing monetary reward in a sequence learning task leads to a reduction in selection errors, as well as a decrease in reaction times, suggesting faster computation at no cost to accuracy (Wachter et al., 2009). Similarly, in a saccade task, reward reduced participant's reaction time whilst making them less sensitive to distractors (Manohar et al., 2015). It has also been shown that reward invigorates movement execution by increasing peak velocity and accuracy during saccades (Manohar et al., 2015; Takikawa et al., 2002) and reaching movements (Carroll et al., 2019; Galaro et al., 2019; Summerside et al., 2018). Therefore, this body of work suggests that reward can consistently shift the speed-accuracy function of both the selection and execution components of a wide range of simple motor behaviours.

As a result, the use of reward has generated much interest as a potential tool to enhance training paradigms for high-performance sports and arts, and rehabilitation procedures for clinical populations such as stroke patients (Goodman et al., 2014; Quattrocchi et al., 2017). However, how reward enhances motor control is still unclear, and future progress in enhancing training and rehabilitation procedures hinges on a more detailed understanding of underlying mechanisms. For instance, in real life, action selection and execution can be intertwined (Chen, Holland & Galea, 2018; Ames et al., 2019; Diedrichsen & Kornysheva, 2015; Orban de Xivry et al., 2017) because many actions are performed continuously, and generalisation between discrete and continuous movements have not always been observed (Ikegami et al., 2010, 2012). Can reward affect both selection and execution concomitantly without interference in reaching movements?

Another open question is how reward mechanically drives improvements in performance. Recent work in eye and reaching movements suggests that reward acts by increasing feedback control, enhancing one's ability to correct for movement error (Carroll et al., 2019; Manohar et al., 2019), which could explain selection improvements as well. However, there are far simpler mechanisms which reward could utilize to improve execution. For example, the motor system has the ability to control the stiffness of its effectors, such as the arm during a reaching task, by employing co-contraction of two antagonist muscles at once (Gribble et al., 2003; Perreault et al., 2002). This increase in arm stiffness results in the limb being more stable in the face of perturbations (Franklin et al., 2007), and capable of absorbing noise that may arise during the movement itself (Selen et al., 2009; Ueyama & Miyashita, 2013), thus reducing error and improving performance (Gribble et al., 2003). Yet, it is unclear whether the reward-based improvements in execution are related to increased arm stiffness.

To address this, we devised a reaching task in which participants could be rewarded with money as a function of their reaction time and movement time. Occasionally, distractor targets of a different colour appeared, and participants were told to withhold movement until the correct target subsequently appeared, allowing for a selection component to be quantified. In a first experiment, we show that reward improves both selection and execution concomitantly, and that the presence or absence of reward, rather than reward magnitude modulated this effect. In a second experiment, we asked whether punishment had a similar effect to reward. We demonstrate that although both reward and punishment led to similar effects, action execution, but not action selection, showed a more global, non-contingent sensitivity to punishment. Behavioural and computational analysis suggested that in addition to an increase in feedback corrections, reward may have improved motor execution through an increase in arm stiffness. In a third and fourth experiment, we tested this hypothesis and provide direct evidence that reward is associated with an increase in arm stiffness. Therefore, reward not only invigorates motor performance by increasing the contribution of feedback control, but also protects against noise at the peripheral level via an increase in arm stiffness.

4.2 Methods

4.2.1 Participants

30 participants (2 males, median age: 19, range: 18-31) took part in experiment 1. 30 participants (4 males, median age: 20.5, range: 18-30) took part in experiment 2. 30 participants (10 male, median age: 19.5, range: 18-32) took part in experiment 3, randomly divided into two groups of 15. 20 participants (2 male, median age: 19, range: 18-20) took part in experiment 4. All participants were recruited on a voluntary basis and were rewarded with money (£7.5/h) or research credits depending on their choice. Participants were all free of visual (including colour discrimination), psychological or motor impairments. The study was approved by and done in accordance with the University of Birmingham Ethics Committee under the project code ERN_09-528P.

4.2.2 Task design

Participants performed the task on an end-point KINARM (BKIN Technologies, Ontario, Canada). They held a robotic handle that could move freely on a plane surface in front of them, with the handle and their hand hidden by a panel (figure 4.1A). The panel included a mirror that reflected a screen above it, and participants performed the task by looking at the reflection of the screen, which appeared at the level of the hidden hand. The sampling rate was 1kHz.

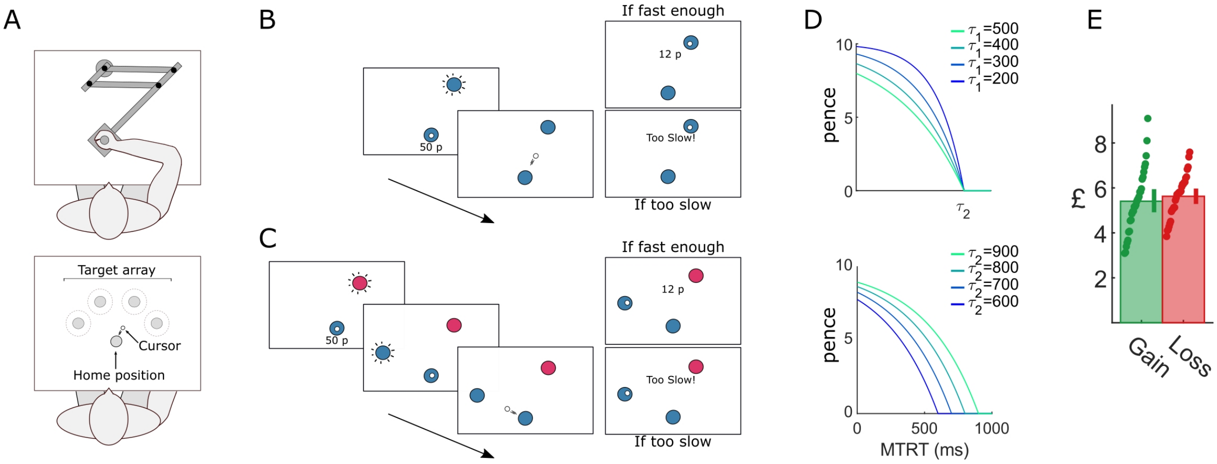


Figure 4.1: Reaching paradigm. A. Participants reached to a series of targets using a robotic manipulandum. B. Normal trial. Participants reached at a single target and earned money based on their performance speed. Speed was the sum of movement time and reaction time (MTRT). If they were too slow (MTRT<τ_2), a message "*Too slow!*" appeared instead of the reward information. C. Distractor trial. Occasionally, a first target bearing a different colour appeared, and participants were told to wait for the second, correct target to appear and reach toward the latter. D. The faster participants moved, the more money they made. The function varied based on two parameters τ_1 and τ_2. The upper and lower plots show how the function varied as a function of τ_1 (τ_2 fixed at 800ms) and τ_2 (τ_1 fixed at 400ms), respectively, for a 10p trial. See methods for more details. E. During the second experiment, participants earned on average the same amount of money during rewarded trials as they lost during punishment trials (see section 4.2.3 for details).

Each trial started with the robot handle bringing participants 4cm ahead of a fixed starting position, except for experiments 3-4 to avoid interference with the perturbations during catch trials. A 1cm diameter starting position then appeared, bearing a colour that matched one of several possible reward values, depending on the experiment. The reward value was also displayed in text under the starting position (figure 4.1B-C). Once participants entered the starting position, a 1cm target appeared 20 cm away from the starting position, and participants were instructed to move as fast as they could towards it and stop in it. They were informed that a combination of their reaction time and movement time defined how much money they would receive, and that this amount accumulated across the experiment. They were also informed that end-position was not factored in as long as they were within 4cm of the target centre. These instructions ensured that participants had a similar approach to the task, since pilot experiments showed some participants put more emphasis or on accuracy.

The reward function was a close-loop design that incorporated the recent history of performance, to ensure that participants received similar amounts of reward, and that the task remained consistently challenging over the experiment (Manohar et al., 2015; Reppert et al., 2018). To that end, the reward function was defined as:

$$r_t = r_{max} \cdot max(1 - e^{(\frac{MTRT-\tau_2}{\tau_1})}, 0) \tag{4.1}$$

where r_{max} was the maximum reward value for a given trial, $MTRT$ the sum of RT and MT, and τ_1 and τ_2 adaptable parameters varying as a function of performance (figure 4.1D). Specifically, τ_1 and τ_2 were the median of the last 20 trials' 3-4th and 16-17th fastest MTRTs, respectively, and were initialised as 400 and 800 at the start of each participant training block. τ values were constrained so that $\tau_1 < \tau_2 < 900$ is always true. In practice, all reward values were rounded up (or down in the punishment condition of experiment 2) to the next penny so that only integer penny values would be displayed. While the main structure of the reward function was taken from Manohar et al. (2015), the parameter values were defined as above following pilot experiments to ensure that reward values obtained by participants were large enough and consistent enough across participants.

Targets were always of the same colour as the starting position (figure 4.1B). However, in experiments 1-2, occasional distractor targets appeared with these being defined by a different colour than the starting position (figure 4.1C). Participants were informed to ignore these targets and wait for the second target to appear. Failure to comply in rewarded and punished trials resulted in no gains for this trial and an increase in loss by a factor of 1.2, respectively. The first target (distractor or not) appeared 500-700ms after entering the starting position using a uniform random distribution, and correct targets in distractor trials appeared 300-600ms after the distractor target using the same distribution.

When reaching movement velocity passed below a 0.3 m/s threshold, end-position was recorded, and monetary gains were indicated at the centre of the workspace. After 500ms, the robotic arm then brought the participant's hand back to the initial position 4cm before the starting position.

In every experiment, participants were first exposed to a training block, where all targets had the same reward value equal to the mean of all value combinations used later in the experiment (*e.g.* if the experiment had 0p and 50p trials, the training reward amounted to 25p per trial). Participants were informed that money obtained during the training will not count toward the final amount they would receive. Starting position and target colours were all grey during training. The τ values obtained at the end of training were then used as initial values for the actual task.

4.2.3 Experimental design

Experiment 1: reward magnitude experiment

The purpose of this experiment was to asses concomitant sensitivity to reward of the selection and execution components for different reward magnitudes. There were 4 possible target locations positioned every 45° around the midline of the workspace, resulting in a 135° span (figure 4.1A). Participants first practiced the task in a 48-trial training block to get acquainted to the apparatus and basic task design. They then experienced a short block (24 trials) with no distractors, and then a main block of 168 trials (72 distractors, 42.86%). The proportion of distractor-containing versus non distractor-containing trials was determined with pilot experiments, so as to increase propensity to get distracted and thus avoid a “ceiling” effect for the selection component. Trials were randomly shuffled within each block. Reward values used during the task were 0, 10 and 50p, similar to Manohar et al. (2015), which allowed to test for different reward magnitudes.

Experiment 2: reward-punishment experiment

In this experiment, the effect of punishment compared to reward on both action selection and execution was assessed, because reward and punishment have previously been shown to lead to dissociable effects in motor learning (Galea et al., 2015). The same 4 target positions were used as experiment 1, and participants first practiced the task in a 48 trials training block. Participants then performed a no-distractor block and a distractor block (12 and 112 trials) in a rewarded condition (0p and 50p trials) and then in a punishment condition (-0p and -50p trials), in a counterbalanced fashion across participants. In the distractor blocks, 48 trials were distractor trials (42.86%). Before the punishment blocks, participants were told that they would receive a starting money pool of £11 and that the slower they moved, the more money they lost. This resulted in participants gaining on average a similar amount of money on the reward and punishment blocks. They were also informed that if they missed the target or went to the distractor target, their losses on that trial would be multiplied by a factor of 1.2. The reward function was biased so that:

$$r_t = -r_{max} \cdot max(1 - e^{(\frac{MTRT - \tau_2 + a}{\tau_1 + b})}, 0) \tag{4.2}$$

With $a = 268.5$ and $b = -71.4$. The update rule was also altered, with τ_1 and τ_2 the median of the last 20 trials’ 15-16th and 17-18th fastest MTRTs, respectively. These new updating indexes and a and b parameters were obtained by using the *lsqnonlin* function of

Matlab's *Optimization toolbox*: the performance data of the reward-magnitude experiment was fitted to a punishment function with free a and b parameters, free starting money pool value and free updating indexes. The *lsqnonlin* function then minimised the difference in average losses compared to the average gains observed in the reward-magnitude experiment. Experimentally, participants gained on average £5.40 in the reward condition and lost £5.63 in the punishment condition (paired t-test: $t(29) = -0.55, p = 0.58, d = -0.1$; figure 4.1E), meaning that this manipulation successfully allowed for a similar amount of gains and losses for a given participant.

Experiment 3: end-of-reach stiffness

In this task, we aimed at measuring end-point stiffness of the arm at the end of the movement—specifically, right after the movement stops. Our hypothesis is that stiffness was increased in a rewarding condition as opposed to a no-reward condition, as this could explain how execution performance can increase while speed also increases. Each of two groups reached to a target located 20cm from the starting position, at +45 and −45° from the midline for the first and second group, respectively. On occasional catch trials, when movement velocity passed under a 0.3 m/s threshold, a 300ms-long, 8mm displacement pushed participants away from their starting position and back, allowing us to measure end-point stiffness (see section 4.2.4 and 4.3.6 and figure 4.12). No distractor trials were employed in this experiment. This type of displacement profile was used based on previous work showing that it can reliably provide end-point stiffness measurements (Franklin et al., 2003; Selen et al., 2009).

Participants performed two training sessions, one with no catch trials (25 trials) and one with 4 catch trials out of 8 trials, in four possible directions from 0 to 270° around the end position to familiarise participants with the displacement. Participants then performed the main block with 64 catch trials out of 200 trials (32%) and 0p and 50p reward values. During the main block, displacements were in 1 of 8 randomly assigned directions from 0-

315° around the end-position (figure 4.12A). We used sessions of 233 trials to ensure session durations remained short, ruling out any effect of fatigue on stiffness as co-contraction is metabolically taxing. To ensure that any measure of stiffness was not due to differences in grip position or a loose finger grip, participant's hands were restricted with a solid plastic piece which held the wrist straight and a reinforced glove that securely strapped the fingers around the handle during the entire task.

Experiment 4: start-of-reach stiffness

In this last experiment, we tested whether similar differences in end-point stiffness existed between rewarding and no-reward trials at the start of the reach. Based on the same *a priori* analysis as for the end-of-reach stiffness experiment—*i.e.* time-time correlation maps, see section 4.3.3—it should be expected that no difference can be observed at the start of the reach. The experiment was essentially the same as experiment 3, except that the catch trials occurred in the start position (figure 4.14A) at the time the target was supposed to appear. To ensure participants remained in the starting position, two different targets (±45° from midline) were used to maintain directional uncertainty. Participants had 24 trials during the no-catch-trial training, 16 trials during the catch-trial training (8 catch trials), and 200 trials during the main block, with 64 (32%) catch trials.

4.2.4 Data analysis

All the analysis code is available on the *Open Science Framework* website, alongside the experimental datasets at `https://osf.io/7as8g/`. Analyses were all made in Matlab (Mathworks, Natick, MA) using custom-made scripts and functions.

Trials were manually classified as distracted or non-distracted (see figure 4.2). Trials that did not include a distractor target—named *no-distractor* trials—were all considered non-distracted. Distracted trials were defined as trials where a distractor target was

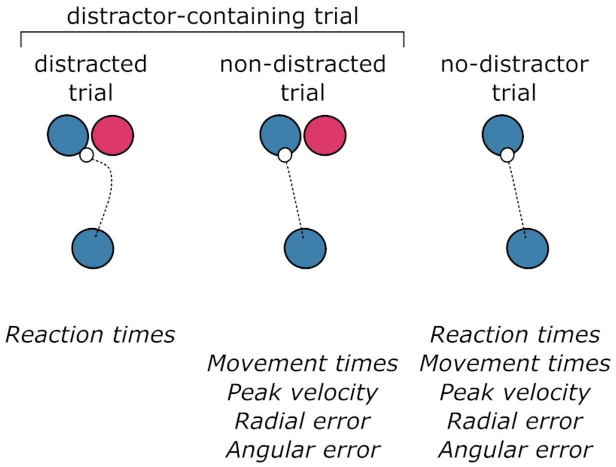


Figure 4.2: Schematic of the different types of trials in the reward-magnitude and reward-punishment experiments and variables that included them in the analysis. Three trial types were distinguished in the experiment. First, trials that included a distracor—named "distractor-containing" trials—were manually classified into "distracted" and "non-distracted" based on the reaching profile of participants. Second, trials that had only a normal target were named "no-distractor" trials. In this schematic, the lower circle represents the starting position, the dashed line the trajectory of a reach, the upper circle of same colour as the starting position was the normal target, and the one with a different colour was a distractor target. The variables indicated in italic under the schematics indicate which trial types were included in their calculations. For instance, movement times were obtained by averaging reaction times of non-distracted and no-distractor trials only and distracted trials were ignored, because they will express a movement duration that is mainly driven by the distracted profile of the reach rather than its speed.

displayed, and participants initiated their movement (*i.e.* exited the starting position) toward the distractor instead of the correct target. If participants readjusted their reach "mid-flight" to the correct target or initiated their movement to the right target and readjusted their reach to the distractor, this was still considered a distracted trial. On very rare occasions (<20 trials in the whole study), participants exited the starting position away from the distractor but before the correct target appeared; these trials were not considered distracted.

Reaction times were measured as the time between the correct target onset and when the participant's distance from the centre of the starting position exceeded 2cm. In trials that were marked as "distracted" (*i.e.* participant initially went to the distractor target), the distractor target onset was used. In distractor-containing trials, the second, correct target did not require any selection process to be made, since the appearance of the distractor target informed participants that the next target would be the right one. For this reason, reaction times were biased toward a faster range in non-distracted trials. Consequently, mean reaction times were obtained by including only no-distractor trials, and distracted trials (figure4.2). For every other summary variable, we included all trials that were not distracted trials, that is, we included non-distracted trials and no-distractor trials (figure4.2).

In experiments 1-2, we removed trials with reaction times higher than 1000ms or less than 200ms, and for non-distracted trials we also removed trials with radial errors higher than 6cm or angular errors higher than 20°. Overall, this resulted in 0.3% and 0.7% trials being removed from experiment 1 and 2, respectively. Speed-accuracy functions were obtained for each participant by binning data in the x-dimension into 50 quantiles and averaging all y-dimension values in a x-dimension sliding window of a 30-centile width (Manohar et al., 2015). Then, each individual speed-accuracy function was averaged by quantile across participants in both the x and y dimension.

Time-time correlation analyses were performed exclusively on non-distracted trials. Trajectories were taken from exiting the starting position to when velocity fell below 0.1m/s. They were rotated so that the target appeared directly in front of the starting position, and y-dimension positions were then linearly interpolated to a hundred evenly spaced timepoints. We focused on the y dimensions because it displays most of the variance (figure 4.3). Correlation values were obtained on y-positions and fisher-transformed before follow-up analyses (Manohar et al., 2019).

For experiments 3-4, positions and servo forces in the x and y dimensions between 140-200ms after perturbation onset were averaged over time for each catch trial (Franklin et al., 2003; Selen et al., 2009). Then, the stiffness values were obtained using multiple linear regressions (function *fitlm* in Matlab). Specifically, for each participant, K_{xx} and K^a_{xy} were the resulting x and y coefficients of $F_x \sim 1 + x + y$ and K^a_{yx} and K_{yy} were the resulting x and y coefficients of $F_y \sim 1 + X + Y$. Then, we can define the asymmetrical stiffness matrix:

$$K_a = \begin{bmatrix} K_{xx} & K^a_{xy} \\ K^a_{yx} & K_{yy} \end{bmatrix} \tag{4.3}$$

And the symmetrical stiffness matrix that we will use in subsequent analysis:

$$K = \begin{bmatrix} K_{xx} & \frac{K^a_{xy}+K^a_{yx}}{2} \\ \frac{K^a_{xy}+K^a_{yx}}{2} & K_{yy} \end{bmatrix} = \begin{bmatrix} K_{xx} & K_{xy} \\ K_{xy} & K_{yy} \end{bmatrix} \tag{4.4}$$

These matrices can be projected in Cartesian space using a sinusoidal transform (see section 4.3.6 for details), resulting in an ellipse. This ellipse can be characterised by its shape, orientation and ratio, which we obtained using a previously described method (Perreault et al., 2002).

4.2.5 Statistical analysis

Although for most experiments we employed mixed-effect linear models to allow for individual intercepts, we used a repeated-measure Analysis of Variance (ANOVA) in experiment 1 to compare each reward magnitudes against each other independently. This allowed us to assess the effect of reward without assuming a magnitude-scaled effect in the first place. Paired-sample t-tests were used when one-way repeated-measure ANOVAs reported significant effects, and effect sizes were obtained using partial η^2 and the Cohen's d method. For experiment 2, we used mixed-effect linear models. For experiments 3 and 4, mixed-effect linear models were also used to account for a possible confound between reward and peak velocity in stiffness regulation, while accounting for individual differences in speed using individual intercepts. Since experiment 3 included a nested design (*i.e.* participants were assigned either to the right or left target but not both), we tested for an interaction using a two-way mixed-effect ANOVA to avoid an artificial inflation of p-values (Zuur, 2009). For all ANOVAs, Bonferroni corrections were applied where appropriate, and post-hoc paired-sample t-tests were used if ANOVAs produced significant results. Bootstrapped 95% CIs of the mean were also obtained and plotted for every group.

Since trials consisted of straight movements toward the target, we considered position in the y dimension—*i.e.* radial distance from the starting position—to obtain time-time correlation maps because it expresses most of the variability. To confirm this, reach trajectories were rotated so the target was always located upfront, and error distribution in the x and y dimension was compared for both experiment 1 (figure 4.3A-B) and 2 (figure 4.3C-D). The y dimension indeed displayed a larger spread in error (experiment 1: $t(11156) = -16.15, p < 0.001$; experiment 2: $t(14852) = -13.68, p < 0.001$). Time-time correlation maps were analysed by fitting a mixed-linear model for each timepoint (Manohar et al., 2019; Zuur, 2009) allowing for individual intercepts using the model $z \sim reward + (1|participant)$, with z the fisher-transformed Pearson coefficient ρ for that timepoint. Then clusters of significance, defined as timepoints with p-values for reward of less than 0.05, were corrected for multiple comparisons using a cluster-wise correction and 10,000 permutations (Maris & Oostenveld, 2007; Nichols & Holmes, 2002). This approach avoids unnecessarily stringent corrections such as Bonferroni correction by taking advantage of the spatial organisation of the time-time correlation maps (Maris & Oostenveld, 2007; Nichols & Holmes, 2002).

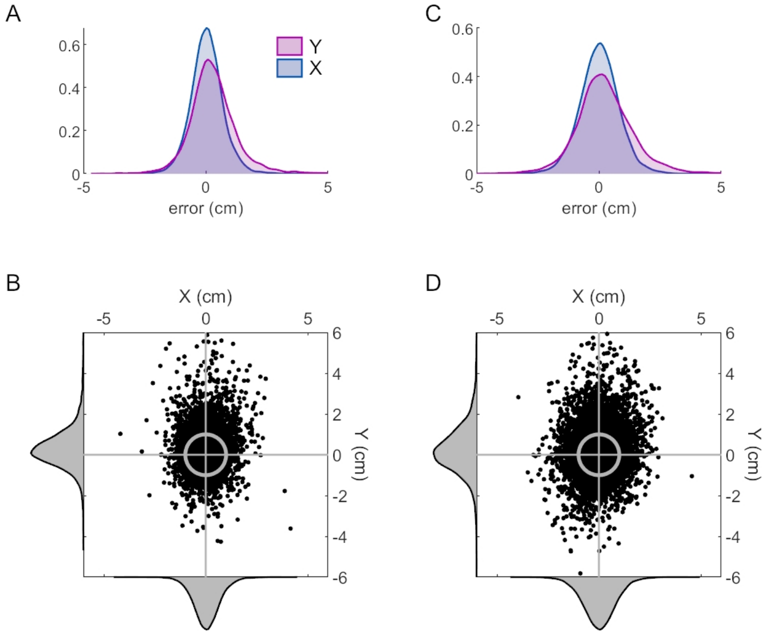


Figure 4.3: Distribution of errors at the end of the reach in the x and y dimension. A. Density function of errors in the x and y dimensions for experiment 1. B. Scatterplot of x versus y error after rotation of all target locations to a frontal location. The horizontal and vertical grey lines indicate the centre of the target, and the circle indicates its size. Density distributions can be observed on the sides. C-D. Same as A-B for experiment 2.

4.2.6 Model simulations

The simulation code is available online on the *Open Science Framework* URL provided above. Simulation results were obtained by running 1000 simulations and obtaining time-time correlation values across those simulations. The sigmoidal step function K used for simulations of the late component was a Gaussian cumulative distribution function such as:

$$K = \frac{1}{\sigma \cdot \sqrt{2\pi}} \int_{-\infty}^{t} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx \tag{4.5}$$

with $\sigma = 0.5, \mu = 0.8$ (or 800 for a 1000 timesteps simulation) and $t_0 < t < t_f$ is the simulation timestep. It should be noted that the use of a sigmoidal function is arbitrary and may be replaced by any other step function, though this will only alter the simulation outcomes quantitatively rather than qualitatively. Values of the feedback control term are taken from Manohar et al. (2019). On the other hand, different noise terms were taken for our simulations because previous work only manipulated one parameter per comparison, whereas we manipulated both noise and feedback at the same time in several models (equations 4.15 and 4.16) and the model is more sensitive to feedback control manipulation than to noise term manipulation.

Regarding model selection, comparisons were performed by fitting each of five datasets to six candidate models:

$$x_{t+1} = x_t + \gamma \cdot \mathcal{N}(\mu, \sigma) \tag{4.6}$$

$$x_{t+1} = x_t + \beta \cdot x_t + \mathcal{N}(\mu, \sigma) \tag{4.7}$$

$$x_{t+1} = x_t - 0.002\, x_t + (1 + \gamma \cdot K) \cdot \mathcal{N}(\mu, \sigma) \tag{4.8}$$

$$x_{t+1} = x_t + (-0.002 + \beta \cdot K) \cdot x_t + \mathcal{N}(\mu, \sigma) \tag{4.9}$$

$$x_{t+1} = x_t + (-0.002 + \beta) \cdot x_t + (1 + \gamma \cdot K) \cdot \mathcal{N}(\mu, \sigma) \tag{4.10}$$

$$x_{t+1} = x_t + (-0.002 + \beta \cdot K) \cdot x_t + (1 + \gamma) \cdot \mathcal{N}(\mu, \sigma) \tag{4.11}$$

with equation 4.6 representing a model with noise reduction, equation 4.7 a model with increased feedback control, equation 4.8 a model with late noise reduction, equation 4.9 a model with late increase in feedback control, equation 4.10 a model with increased feedback and late noise reduction and equation 4.11 a model with late noise reduction and increased feedback. The free parameters were β and γ, with the last two model including both of them and all the others including one, according to the equations. K was a step function as indicated in equation 4.5 and was fixed. 1000 simulations were done with 100 timesteps per simulation. Time-time correlation maps were then fisher-transformed and substracted to a control model $x_{t+1} = x_t + \mathcal{N}(\mu, \sigma)$ for equation 4.6 and $x_{t+1} = x_t - 0.002 \cdot x_t + \mathcal{N}(\mu, \sigma)$ for all other models to obtain contrast maps. The resulting contrast maps were then fitted to the empirical contrast maps obtained to minimise the sums of squared errors for each individual for individual-level analysis, and across individuals for the group-level analysis. Of note, rather than fitting the model to the across-participant averaged contrast map in the group-level analysis, the model minimised all the individual maps at once, allowing for a single model fit for the group without averaging away individual map features. The optimisation process was done using the *fminsearch* function of the *Optimization* toolbox in Matlab. The free parameter search was initialised with $\beta_0 = 0$ and $\gamma_0 = 0$. Model comparisons were performed by finding the model with lowest Bayesian information criterion (BIC), defined as $BIC = n\log(RSS/n) + k\log n$ with $n = 100^2 = 10000$ the number of timepoint per participant map, k the number of parameters in the model considered and RSS the model's residual sum of squares.

4.3 Results

4.3.1 The effect of reward magnitude on action selection and action execution

Experiment 1 examined the effect of reward on the selection and execution components of a reaching movement. Whilst holding a robotic manipulandum, participants (N=30) made discrete reaching movements towards 1 of 4 visual targets presented 20cm away from a central start position (figure 4.1A). To assess the effect of reward value on reaching performance, participants were informed of the upcoming trial type prior to movement onset: 0p, 10p and 50p. For the 10p and 50p trials participants could earn money based on their combined reaction time and movement time. The scoring function which translated performance to monetary gain was adaptive (figure 4.1D), factoring in the recent history of movement times and reaction times to ensure participants experienced comparable amounts of reward despite idiosyncrasies in individual's reaction times and movement speed (Berret et al., 2018; Reppert et al., 2018). To assess selection and execution performance concomitantly, we interleaved normal trials and distractor-containing trials. In normal trials, the target's colour matched the starting position colour (figure 4.1B), while in distractor-containing trials (42% of trial) a distractor target bearing a different colour than the starting position appeared prior to the correct target (figure 4.1C). In this case, participants were instructed to withhold their movement to the distractor and wait until the correct target appeared before making a movement. If participants exited the starting position upon appearance of a distractor, the trial was considered as "distracted" (*e.g.* figure 4.2). While one's propensity to initiate reaches to a distractor target provided a measure of selection accuracy, the associated reaction times provided a selection speed, allowing us to define a speed-accuracy function (Fitts, 1954; Hübner & Schlösser, 2010; Manohar et al., 2015). For execution, radial error provided a measure of execution accuracy while peak velocity during the reach and movement time provided an execution speed, again allowing us to define a speed-accuracy function.

Participants showed a clear and consistent improvement in selection accuracy in the presence of reward. Specifically, they were less likely to be distracted in rewarded trials, though this was independent of reward magnitude (repeated-measures ANOVA, $F(2) = 15.8, p < 0.001$, partial $\eta^2 = 0.35$, *post-hoc* 0p vs 10p $t(29) = -3.34, p = 0.005, d = -0.61$; 0p vs 50p $t(29) = -5.32, p < 0.001, d = -0.97$; 10p vs 50p $t(29) = -2.21, p = 0.07, d = -0.49$; figure 4.4A). However, this did not come at the cost of slowed decision-making, as reaction times remained largely similar across reward values; if anything, reaction times were slightly shorter if a large reward (50p) was available compared to no-reward (0p) trials, though this was not statistically significant ($F(2) = 2.35, p = 0.10$, partial $\eta^2 = 0.07$; figure 4.4B-C).

In addition, reward led to a strong improvement in action execution across participants. Specifically, peak velocity drastically increased with reward value ($F(2) = 43.0, p < 0.001$, partial $\eta^2 = 0.60$, *post-hoc* 0p vs 10p $t(29) = -7.40, p < 0.001, d = -1.35$; 0p vs 50p $t(29) = -7.61, p < 0.001, d = -1.39$; 10p vs 50p $t(29) = -3.52, p = 0.003, d = -0.64$; figure 4.4D). Unsurprisingly, movement time also showed a similar effect, that

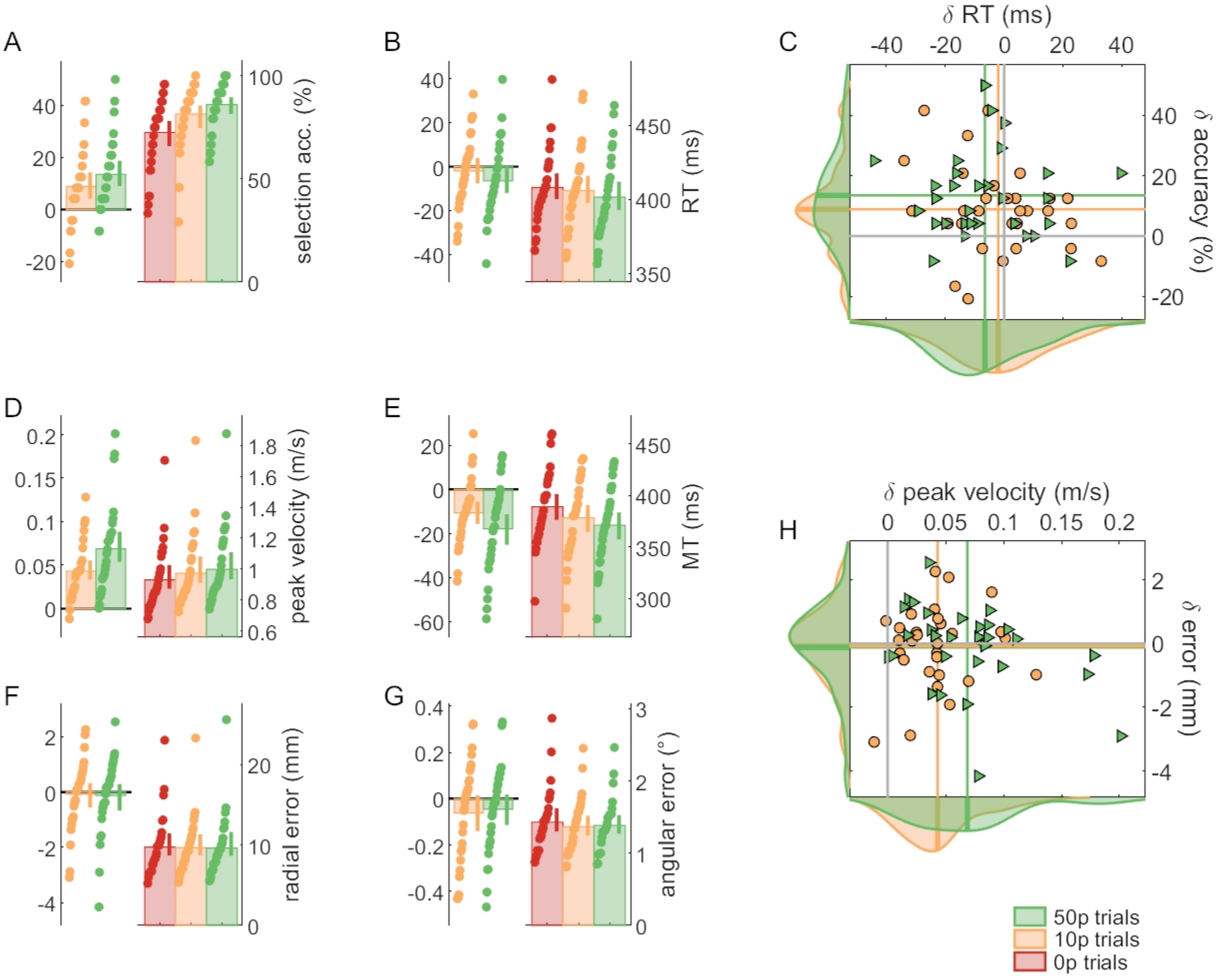


Figure 4.4: Reward enhances performance in both selection and execution. For all bar plots, means of summary variables are represented for each trial type (0p, 10p, 50p) on the right-hand side, and data normalised to 0p performance for each participant are displayed on the left-hand side. Dots represent individual values and error bars indicate bootstrapped 95% CIs of the mean. A. Selection accuracy, as the percentage of trials where participants initiated reaches toward the correct target instead of the distractor target. B. Mean reaction times. C. Scatterplot of mean reaction time against selection accuracy. Values are normalised to 0p trials. The coloured lines indicate the mean value for each condition, and the solid grey lines indicate the origin, that is, 0p performance. Data distributions are displayed on the sides, with transversal bars indicating the mean of the distribution. D. Mean peak velocity during reaches. E. Mean movement times of reaches. F. Mean radial error at the end of the reach. G. Mean angular error at the end of the reach. H. Scatterplot showing execution speed (peak velocity) against execution accuracy (radial error), similar to C.

is, mean movement time decreased with reward, though this did not scale with reward magnitude ($F(2) = 15.3, p < 0.001$, partial $\eta^2 = 0.35$, *post-hoc* 0p vs 10p $t(29) = 4.07, p < 0.001, d = 0.74$; 0p vs 50p $t(29) = 4.99, p < 0.001, d = 0.91$; 10p vs 50p $t(29) = 2.08, p = 0.09, d = 0.38$; figure 4.4E). However, this reward-based improvement in speed did not come at the cost of accuracy as radial error ($F(2) = 0.15, p = 0.86$, partial $\eta^2 = 0.005$) and angular error ($F(2) = 1.51, p = 0.23$, partial $\eta^2 = 0.05$) remained unchanged (figure 4.4F-H). Finally, we analysed how speed-accuracy functions were altered by reward (figure 4.5). To this end, trials for each reward value and participant were sorted as a function of their speed (reaction time for selection and peak velocity for execution) and divided into 50 quantiles (Manohar et al., 2015). For each quantile, the average accuracy (percentage of non-distracted trials and radial error) over a 30% centile window was obtained. Group averages were then obtained for each quantile in the speed and accuracy dimension, and results are displayed in figure 4.5. As expected, reward shifted the speed-accuracy functions for both selection and execution, underlining augmented motor performance with reward.

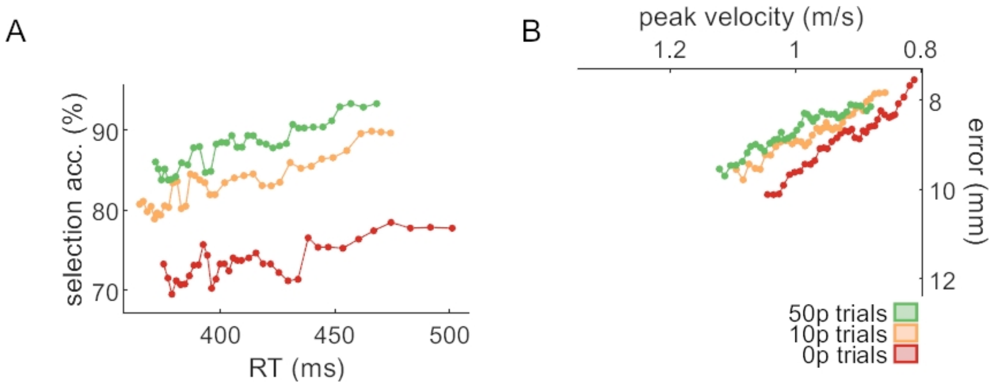


Figure 4.5: Speed-accuracy functions for selection (A) and execution (B) shift as reward values increase. The functions are obtained by sliding a 30% centile window over 50 quantile-based bins. A. For the selection panel, the count of non-distracted trials and distracted trials for each bin was obtained, and the ratio (100*non-distracted/total) calculated afterwards. B. For the execution component, the axes were inverted to match the selection panel in A, *i.e.* the upper left corner indicates faster and more accurate performance. See methods section 4.2.4 and text for details.

This demonstrates that reward enhances the selection and execution components of a reaching movement simultaneously and without interference. However, reward magnitude had only a marginal impact on the effect of reward itself, as opposed to the presence or absence of reward *per se*. Consequently, for the remaining studies, we used the 0p and 50p trial conditions to assess the impact of reward on reaching performance.

4.3.2 The effect of punishment on action selection and action execution

Next, we asked if punishment led to the same effect as reward, as previous reports have shown that they have dissociable effects on the motor system (Galea et al., 2015; Hamel et al., 2018; Song & Smiley-Oyen, 2017; Wachter et al., 2009). A new group of participants (N=30) experienced a reward and a punishment block in a counterbalanced order. In the reward block, 0p and 50p trials were randomly interleaved. Similar to the previous experiment, on 50p trials participants received money as a result of fast reaction times and movement times. The punishment block consisted of randomly interleaved -0p and -50p trials which indicated the maximum amount of money that could be lost on a single trial. At the beginning of this block, participants were given £11, and on -50p trials, participants lost money as a result of slow reaction times and movement times.

To examine these results, we fitted a mixed-effect linear model $DV \sim 1 + RP + value + RP * value + (1|participant)$ that included individual intercepts and an interaction term, where DV is the dependent variable considered, RP indicated whether the context was reward or punishment (*i.e.* reward block or punishment block) and $value$ indicated whether the trial is a baseline trial bearing no value (0p and -0p) or a rewarded/punished trial bearing high value (50p and -50p). Once again value improved selection accuracy ($\beta = 9.72$, CI= $[4.51, 14.9]$, $t(116) = 3.70, p < 0.001$; figure 4.6A) without any effect on reaction times ($\beta = -0.007$, CI= $[-0.015, 0.002]$, $t(116) = -1.53, p = 0.13$; figure 4.6B,C) and increased peak velocity and movement time (main effect of value on peak velocity $\beta = 0.096$, CI= $[0.045, 0.147]$, $t(116) = 3.76, p < 0.001$; on movement time

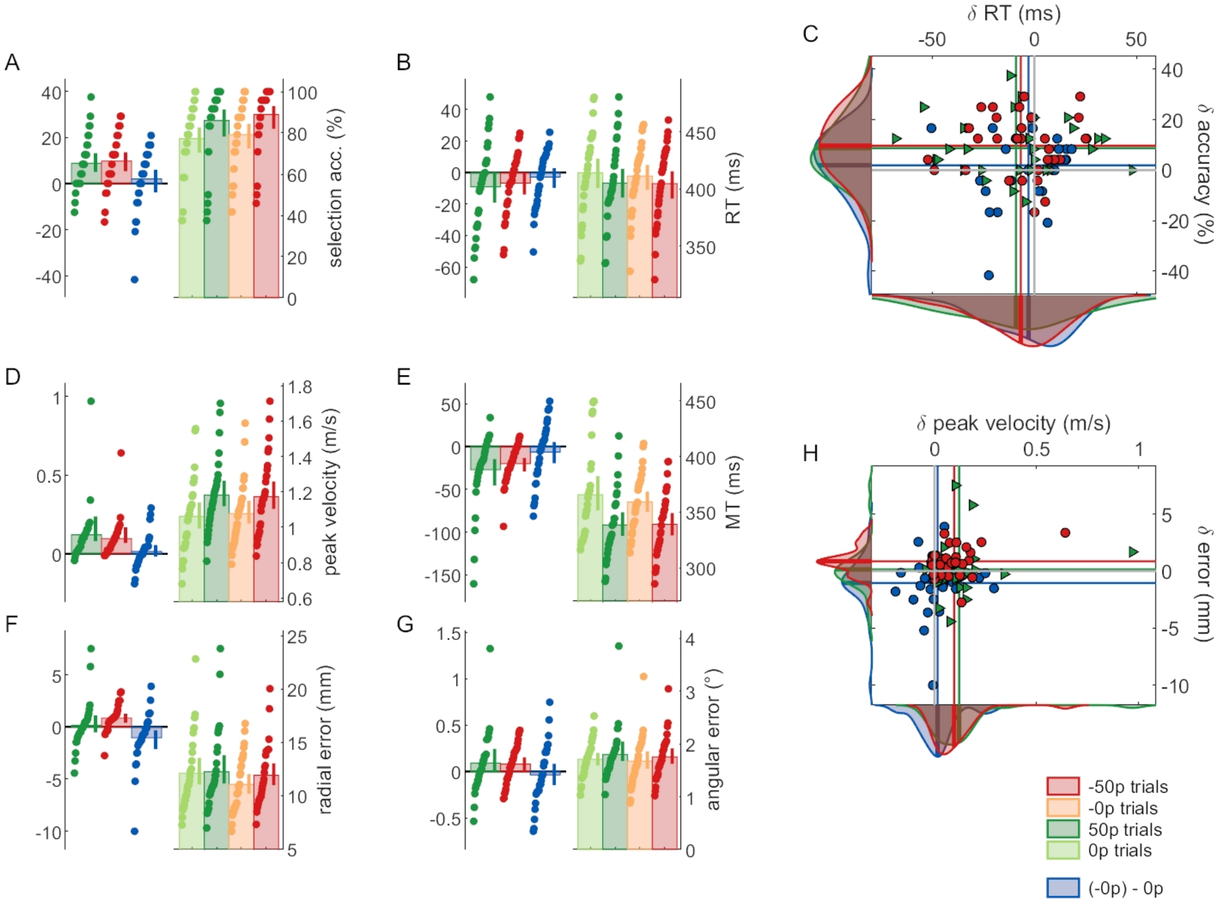


Figure 4.6: Reward and punishment have a similar effect on selection, but not on execution. For all bar plots, summary variables are represented for each trial type (0p, 50p, -0p, -50p) on the right-hand side, and data normalised to baseline performance (0p or -0p depending on the block) for each participant are displayed on the left-hand side, alongside baseline differences. Dots represent individual values, bars indicate the mean value and error bars indicate bootstrapped 95% CIs of the mean. A. Selection accuracy. B. Mean reaction times for each participant. C. Scatterplot of mean reaction time against selection accuracy. Values are normalised to 0p trials. The coloured lines indicate mean values for each condition, and the solid grey lines indicate the origin, that is, 0p performance. Data distributions are displayed on the sides, with transversal bars indicating the mean of the distribution. D. Mean peak velocity. E. Movement times. F. For radial error, punishment did not protect against an increase in error, while reward did. However, a difference can be observed between the baselines (blue bar). G. Angular error. H. Scatterplot showing execution speed (peak velocity) against execution accuracy (radial error), similar to C.

$\beta = -0.02$, CI= $[-0.033, -0.007]$, $t(116) = -3.15, p = 0.002$; figure 4.6D,E) at no accuracy cost (radial error $\beta = -0.085$, CI= $[-0.001, 0.171]$, $t(116) = 1.96, p = 0.052$; angular error $\beta = 0.081$, CI= $[-0.027, 0.189]$, $t(116) = 1.49, p = 0.14$; figure 4.6F-H), therefore replicating the findings from experiment 1. Importantly, context (reward vs. punishment) did not alter these effects on selection accuracy (main effect of block $\beta = -1.94$, CI= $[-7.15, 3.26]$, $t(116) = -0.74, p = 0.46$; interaction $\beta = -0.97$, CI= $[-8.34, 6.39]$, $t(116) = -0.26, p = 0.79$; figure 4.6A), reaction times (main effect of block $\beta = -0.003$, CI= $[-0.006, 0.011]$, $t(116) = -0.66, p = 0.51$; interaction $\beta = -0.002$, CI= $[-0.014, 0.010]$, $t(116) = -0.38, p = 0.70$; figure 4.6B) and peak velocity (main effect of block $\beta = -0.015$, CI= $[-0.066, 0.036]$, $t(116) = -0.59, p = 0.56$; interaction $\beta = -0.024$, CI= $[-0.047, 0.096]$, $t(116) = -0.67, p = 0.50$; figure 4.6D). Interestingly, however, the punishment context did affect radial accuracy, with accuracy increasing compared to the rewarding context (main effect of block, $\beta = 0.10$, CI= $[0.019, 0.19]$, $t(116) = 2.42, p = 0.017$; figure 4.6F), although no interaction was observed ($\beta = -0.07$, CI= $[-0.19, 0.05]$, $t(116) = -1.16, p = 0.25$). Based on the comparison between the baselines (+0p and -0p; blue bar in figure 4.6F), it appears that this effect or reduced radial error is indeed at least partially driven by the baseline trials during the punishment context. This tends to suggest a non-contingent effect of punishment, while reward seems to affect execution accuracy on contingent trials only.

Next, we obtained speed-accuracy functions for the selection and execution components in the same way as for experiment 1 (figure 4.7). While punishment had a similar effect on selection (Figure 4.7A), it produced dissociable effects on execution (Figure 4.7B). Specifically, while peak velocity increased with punishment similarly to reward, it was accompanied by an increase in radial error. Although this could suggest that punishment does not cause a change in the speed-accuracy function relative to its own baseline (-0p) trials, an important result to highlight is that a clear shift in the speed-accuracy function could be seen between the baseline trials of the reward and punishment conditions (Figure 4.7B). Therefore, relative to reward, a punishment context indeed appeared to have a non-contingent beneficial effect on motor execution, in line with the observation on figure 4.6F. Of note, the reward (+50p) condition led to a weaker shift of the speed-accuracy function from the no-reward (+0p) condition compared to what was observed on the first experiment. This effect was mainly observed at higher speeds. It is possible that, at the group level, the presence of the punishment block reduced the motivational saliency of reward in the subsequent block, that is, it led to an interaction effect that decreased the effect size of reward. However, it can be seen from figure 4.6, that the effect of reward remains particularly strong at the within-participant level.

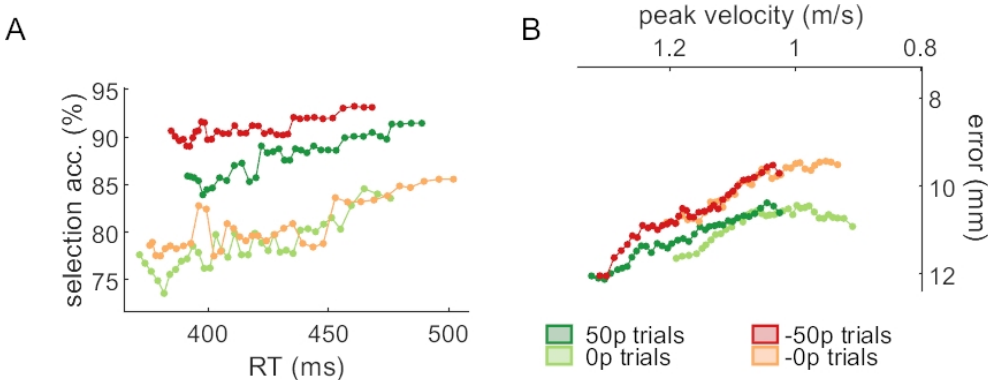


Figure 4.7: Reward and punishment speed-accuracy functions for selection (A) and execution (B) components. The functions are obtained by sliding a 30% centile window over 50 quantile-based bins. A. For the selection panel, the count of non-distracted trials and distracted trials for each bin was obtained, and the ratio (100*non-distracted/total) calculated afterwards. B. For the execution component, the axes were inverted to match the selection panel in A, *i.e.* the upper left corner indicate faster and more accurate performance. See methods section 4.2.4 and text for details.

4.3.3 Time-time correlation analysis

How do reward and punishment lead to these improvements in motor performance? In saccades, it has been suggested that reward increases feedback control, allowing for more accurate end-point performance. To test for this possibility, we performed the same time-time correlation analysis as described in Manohar et al. (2019). Specifically, we assessed how much the set of positions at time t across all trials correlated with the set of positions at time $t + 1$. If movements are stereotyped across trials, this correlation will be high because the early position will provide a large amount of information about the later position. On the other hand, if movements are variable, the correlation will decrease because there will be no consistency in the evolution of position over time. Importantly, the latter occurs with high online feedback because corrections are not stereotyped, but

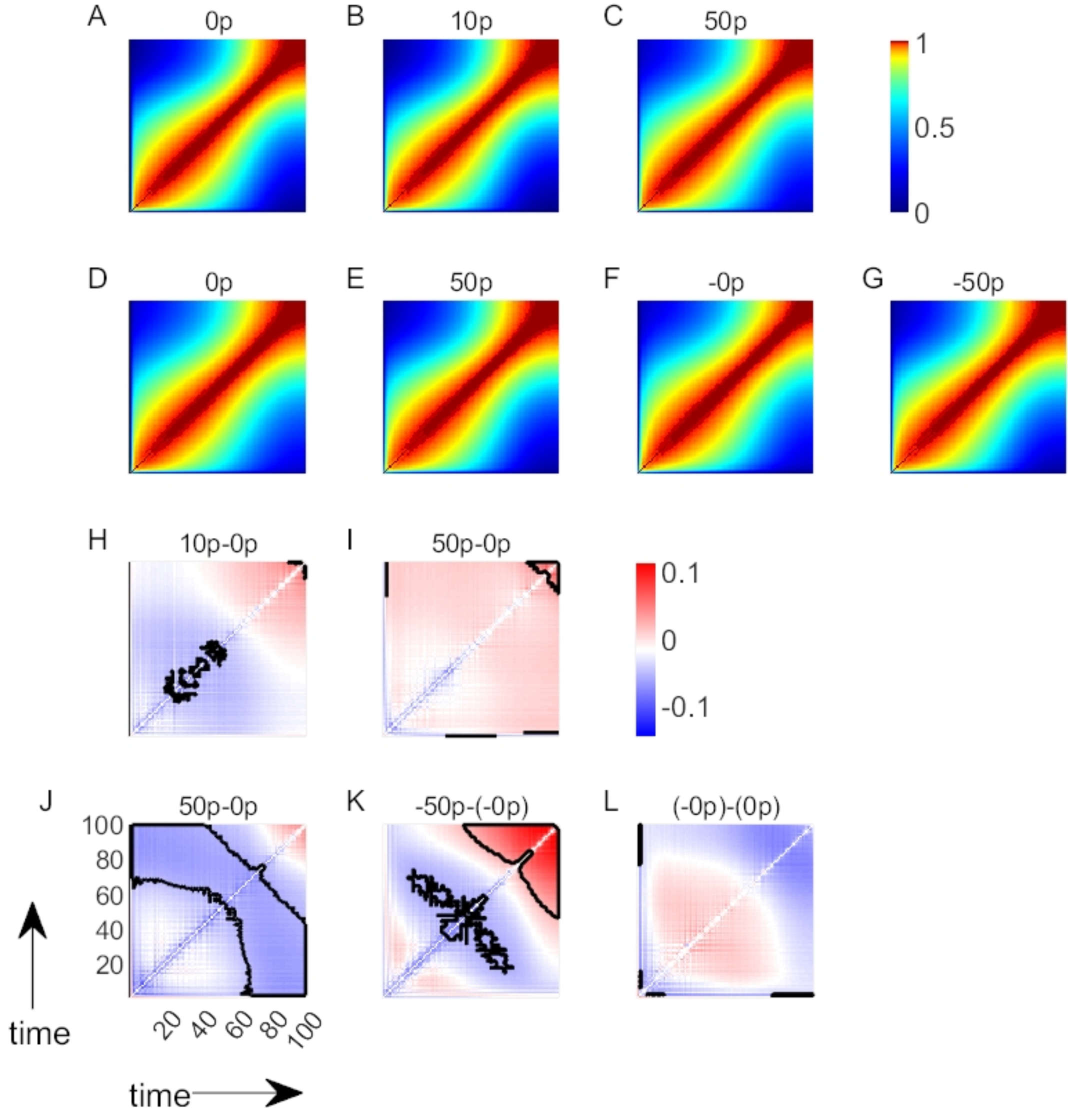


Figure 4.8: Time-time correlation maps show that monetary reward and punishment have a biphasic effect on the reach timecourse. A-C. Time-time correlation maps for all trial types (0p, 10p 50p) in Experiment 1. Colours represent Pearson correlation values. For each map, the lower left and upper right corners represent the start and the end of the reaching movement, respectively. Note that the colour maps are non-linear to enhance readability. D-G. Time-time correlation maps for all trial types (0p,50p,-0p,-50p) in Experiment 2. H-I. Comparison of fisher-transformed correlation maps with the respective baseline map (A) for Experiment 1. Clusters of significance after cluster-wise correction for multiple comparisons are indicated by a solid black line. J-L. Similar comparisons for Experiment 2, with each condition's respective baseline (D and F).

rather dependent on the random error on a given trial (Manohar et al., 2019). If the same mechanism is at play during reaching movements than in saccades, a similar decrease in time-time correlations should be observed.

All timepoints correlations were performed by comparing position over trials by centiles, leading to 100 timepoints along the trajectory (figure 4.8A-G). Across experiment 1 and 2, we observed an increase in time-time correlation in the late part of movement with reward/punishment (figure 4.8H-K). In contrast, the early to middle part of movement showed a clear decorrelation. No difference was observed when comparing baseline trials from experiment 2 (figure 4.8L). Surprisingly, this consistent biphasic pattern across conditions and experiments is the opposite to the one observed in saccades (Manohar et al., 2019). Therefore, this analysis suggests that reward/punishment causes a decrease in feedback control during the late part of reaching movements. However, a reduction in feedback control should result in a decrease in accuracy which was not observed in our data. This suggests that another mechanism is being implemented that enables movements to be performed with enhanced precision under reward and punishment.

One possible candidate is muscle co-contraction. By simultaneously contracting agonist and antagonist muscles around a given joint, the nervous system is able to regulate the stiffness of that joint. Although this is an extremely energy inefficient mechanism, it has been repeatedly shown that it is very effective at improving arm stability in the face of unstable environments such as force fields (Franklin et al., 2003). Critically, it is also capable of dampening noise (Selen et al., 2009), which arises with faster reaching movements, and therefore enables more accurate performance (Todorov, 2005). Therefore, it is possible that increased muscle stiffness could, at least partially, underlie the effects of reward and punishment on motor performance.

4.3.4 Simulation of time-time correlation maps with a simplified dynamical system

To assess if the correlation maps we observed are in line with this interpretation, we performed simulations using a simplified control system (Manohar et al., 2019) and evaluate how it responds to hypothesised manipulations of the control system. Let us represent the reach as a discretised dynamical system (Todorov, 2004):

$$x_{t+1} = \alpha \cdot x_t + \beta \cdot u_t + \mathcal{N}(\mu, \sigma) \tag{4.12}$$

The state of the system at time t is represented as x_t, the motor command as u_t, and the system is susceptible to a random gaussian process with mean $\mu = 0$ and variance $\sigma = 1$. α and β represent the environment dynamics and control parameter, respectively. For simplicity, we assume that $\alpha = 1, \beta = 0$ and that $x_0 = 0$. The goal of the system is to maintain the state at 0 for the duration of the simulation. This is equivalent to assuming that x represents error over time and the controller has perfect knowledge of the optimal movement to be performed. As $\alpha = 1$ and $\beta = 0$, any deviation from the optimal movement is solely due to the noise term that contaminates the system at every time step.

We performed 1000 simulations, each including 1000 timesteps, and show the time-time correlation maps of the different controllers under consideration. First, we assume that no feedback has taken place ($\beta = 0$, equation 4.12). The system is therefore only driven by the noise term (figure 4.9A). The controller can reduce the amount of noise, *e.g.* through an increase in stiffness (Selen et al., 2009). This can be represented as $x_{t+1} = x_t + \gamma \cdot \mathcal{N}(\mu, \sigma)$ with $\gamma = 0.5$. However, this would not alter the correlation map (figure 4.9B-C) as was previously shown (Manohar et al., 2019) because the noise reduction occurs uniformly over time. Now, if a feedback term is introduced with $\beta = -0.002$ and $u_t = x_t$, the system includes a control term that will counter the noise and becomes:

$$x_{t+1} = x_t - 0.002 \cdot x_t + \mathcal{N}(\mu, \sigma) \tag{4.13}$$

Higher feedback control ($\beta = -0.003$) would reduce the noise even further. Comparing this high feedback model with the low feedback model (equation 4.13; figure 4.9D-E),

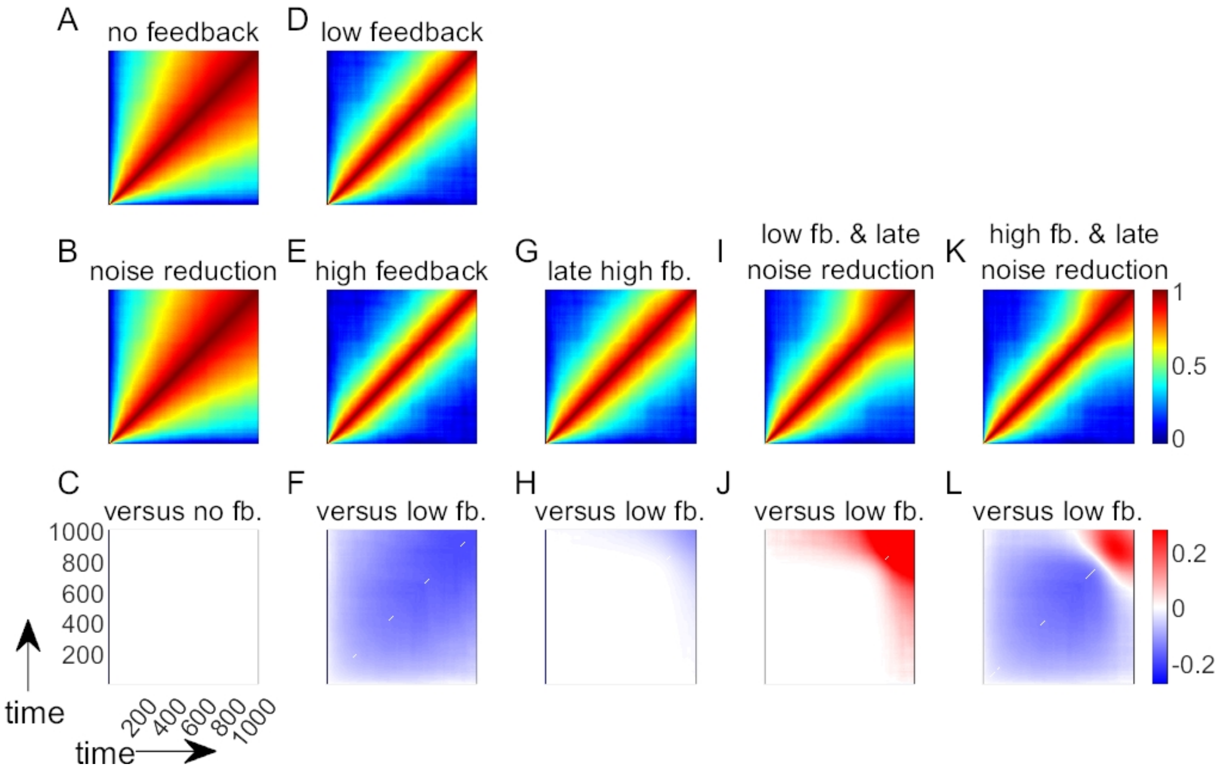


Figure 4.9: Simulations of time-time correlation map behaviour under different models of the reward- and punishment-based effects on motor execution. A,D. Time-time correlation maps of both control models. Colours represent Pearson correlation values. For each map, the lower left and upper right corners represent the start and the end of the reaching movement, respectively. B,E,G,I,K. Time-time correlation maps of plausible alternative models. C,F,H,J,L. Comparison of models with their respective baseline models.

we see that the contrast (figure 4.9F) shows a reduction in time-time correlations similar to what is observed in the late part of saccades (Manohar et al., 2019) and in the early part of arm reaches in our dataset (figure 4.8H-K). Since our dataset displays a biphasic correlation map, it is likely that two phenomena occur at different timepoints during the reach. To simulate this, we altered the original model by including a sigmoidal step function K that is inactive early on ($K = 0$) and becomes active ($K = 1$) during the late part of the reach (see section 4.2.6 for details). This leads to two possible mechanisms, namely, a late increase in feedback or a late reduction in noise:

$$x_{t+1} = x_t + (-0.002 + \beta \cdot K) \cdot x_t + \mathcal{N}(\mu, \sigma) \qquad \beta = -0.001 \tag{4.14}$$

$$x_{t+1} = x_t - 0.002 \cdot x_t + (1 + \gamma \cdot K) \cdot \mathcal{N}(\mu, \sigma) \qquad \gamma = -0.5 \tag{4.15}$$

The results show that a late increase in feedback causes decorrelation at the end of movement (equation 4.14; figure 4.9G-H), which is the opposite of what we observe in our results. However, similar to our behavioural results, a late reduction in noise causes an increase in the correlation values at the end of movement (equation 4.15; figure 4.9I-J). Therefore, our results (figure 4.8H-K) appear to be qualitatively similar to a combined model in which reward and punishment cause a global increase in feedback control and a late reduction in noise (equation 4.16; figure 4.9K-L):

$$x_{t+1} = x_t - 0.003 \cdot x_t + (1 - 0.5 \cdot K) \cdot \mathcal{N}(\mu, \sigma) \tag{4.16}$$

4.3.5 Quantitative model comparison

To formally test which candidate model best describes our empirical observations, we fitted each of them to the experimental datasets. Each of the five empirical conditions displayed in figure 4.8H-L was kept separate, each condition representing a cohort, and their fit assessed separately. While individually fitted models present several advantages over group-level analysis, it has been argued that the most reliable approach to determine the best-fit model is to assess its performance both on individual and group data and compare the outcomes (A. L. Cohen et al., 2008; Lewandowsky & Farrell, 2011) and we will therefore follow this approach. We included six candidate models in our analysis: noise reduction (one free parameter γ; figure 4.9C), increased feedback (one parameter β; figure 4.9E), late feedback (one parameter β; figure 4.9H), late noise reduction (one parameter γ; figure 4.9J), increased feedback with late noise reduction (two parameters β and γ; figure 4.9L) and an additional model with noise reduction and a late increase in feedback control (two parameters β and γ).

Individual-level analysis resulted in the increased feedback with late noise reduction model

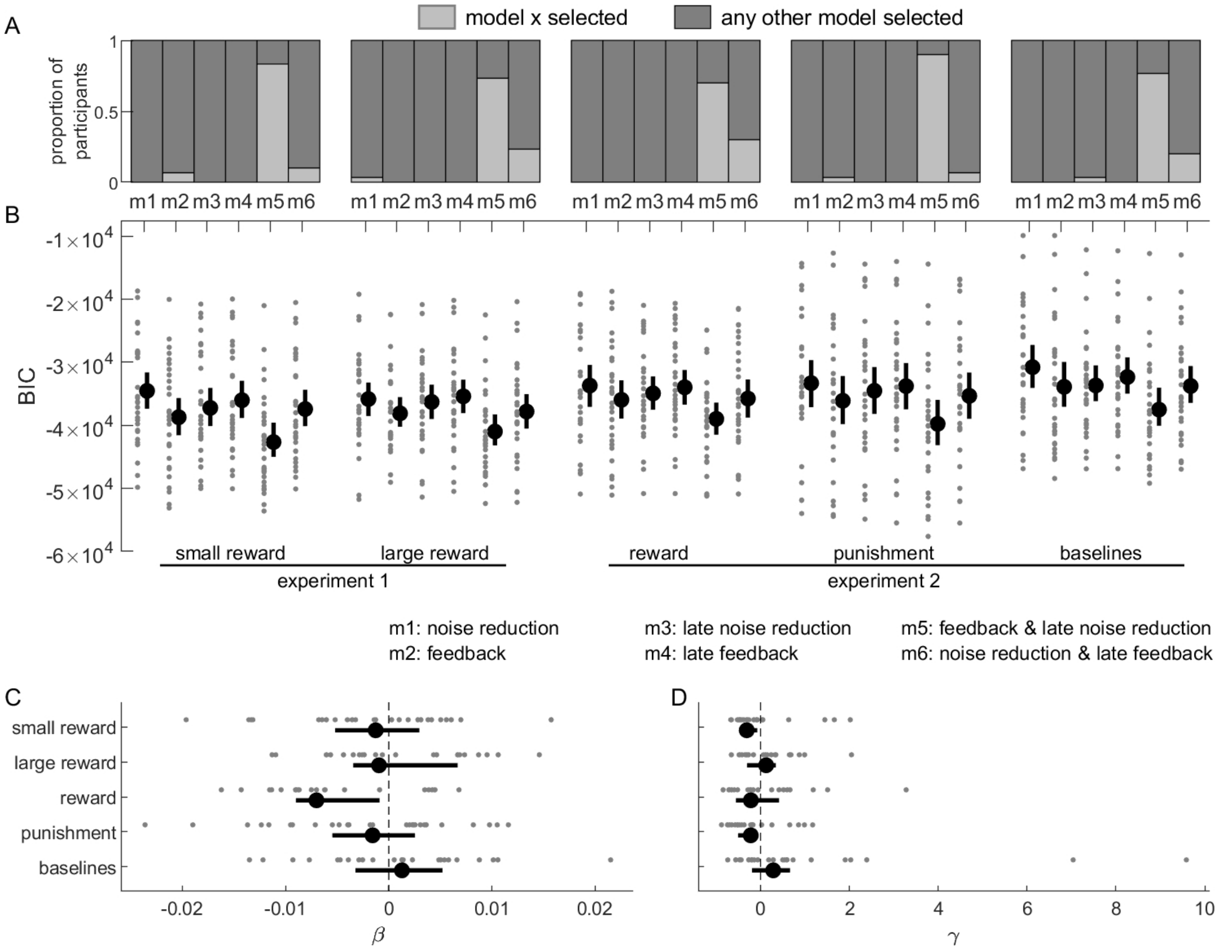


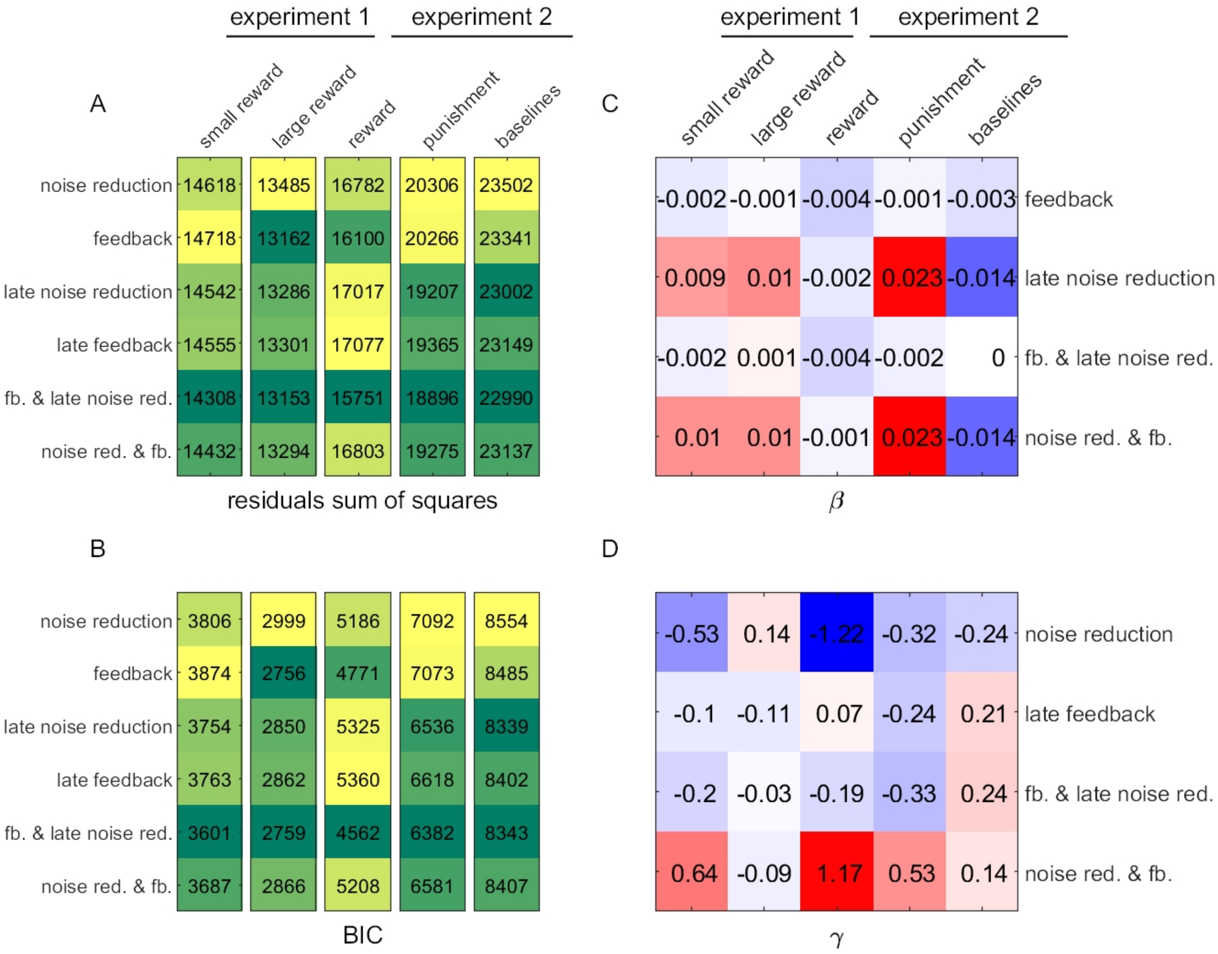
Figure 4.10: Model comparisons for individual fits. A. Proportion of participants whose winning model was the one considered (light gray) against all other models (dark gray) for every cohort. B. Individual and mean BIC values for each participant and each model. Lower BIC values indicate a better fit. Dots indicate individual BICs, the black dot indicates the group mean and the error bars indicate the bootstrapped 95% CIs of the mean. C. β parameter estimate of the model m5 for all participants for whom it was the winning model. The black dots indicate median values and the error bars indicate bootstrapped 95% CIs of the median. D. Same as C for the γ parameter. BIC: Bayesian information criterion.

being selected by a strong majority of participants for each cohort (cohort 1-5: $\chi^2 = [97.6, 76.8, 74.4, 116.8, 83.2]$, all $p < 0.001$, figure 4.10A), confirming qualitative predictions. The best-fit model for each participant was defined as the model bearing the lowest BIC (4.10B). This allowed us to account for each model's complexity, because the BIC penalises models with more free parameters such as the last two models in our set that include both a β and γ parameter. Next, we assessed the best-fit model's estimated parameters value for each participant, excluding those whose best-fit model was not the winning one

(N = [5/30, 8/30, 9/30, 3/30, 7/30] for cohort 1-5, respectively). Individual estimates are displayed in figure 4.10C-D. It would be expected that both β and γ take negatives values, indicating an increase in feedback control and a reduction in noise, respectively. Accordingly, the distribution of parameter estimates shows this trend: the median β parameter was negative for all cohorts except the "baselines" cohort (figure 4.10C), indicating an increase in feedback control; and the median γ was also negative in three out of five cohorts (figure 4.10D), signaling a late reduction in noise. Of note, the "baselines" cohort showed effects opposite to the general trend and shows several runaway parameter estimates. However, this should not be surprising, considering that this cohort is the only one that showed no significant trend in its contrast map (figure 4.8L), and accordingly, it was the cohort displaying the highest BIC for all models considered. Second, though the median value of both parameters were in line with theoretical expectations, the underlying distributions also displayed a large amount of variability, indicating uncertainty regarding the true value of those estimates. This is usually due to high inter-individual variability in the behaviour of interest, or an overly small amount of sampling data per participant. Therefore, those median parameter estimates should be considered with some caution.

Of note, the γ parameter displays some assymetry toward high positive values. This is not surprising because this parameter is *de facto* constrained to $[-1 \ +\infty[$ as for $\gamma = -1$ the noise term $(1 + \gamma) \cdot \mathcal{N}(\mu, \sigma)$ would reduce to 0, making each trial deterministic and the time-time correlations undefined.

To confirm that the selected model is indeed the most parsimonious choice, we compared the individual-level outcome to a group-level outcome. Each candidate model was fit to all individual correlation maps at once, thereby allowing for each free parameter to take a single value per cohort. This is equivalent to assuming that the parameters are not random but rather fixed effects, allowing us to observe the population-level trend with higher certainty, though at the cost of ignoring its variability (A. L. Cohen et al., 2008; Lewandowsky & Farrell, 2011). Again, for every cohort, the model with lowest



	experiment 1		experiment 2		
	small reward	large reward	reward	punishment	baselines
noise reduction	14618	13485	16782	20306	23502
feedback	14718	13162	16100	20266	23341
late noise reduction	14542	13286	17017	19207	23002
late feedback	14555	13301	17077	19365	23149
fb. & late noise red.	14308	13153	15751	18896	22990
noise red. & fb.	14432	13294	16803	19275	23137

	experiment 1		experiment 2		
	small reward	large reward	reward	punishment	baselines
noise reduction	3806	2999	5186	7092	8554
feedback	3874	2756	4771	7073	8485
late noise reduction	3754	2850	5325	6536	8339
late feedback	3763	2862	5360	6618	8402
fb. & late noise red.	3601	2759	4562	6382	8343
noise red. & fb.	3687	2866	5208	6581	8407

experiment 1		experiment 2			
small reward	large reward	reward	punishment	baselines	
-0.002	-0.001	-0.004	-0.001	-0.003	feedback
0.009	0.01	-0.002	0.023	-0.014	late noise reduction
-0.002	0.001	-0.004	-0.002	0	fb. & late noise red.
0.01	0.01	-0.001	0.023	-0.014	noise red. & fb.

experiment 1		experiment 2			
small reward	large reward	reward	punishment	baselines	
-0.53	0.14	-1.22	-0.32	-0.24	noise reduction
-0.1	-0.11	0.07	-0.24	0.21	late feedback
-0.2	-0.03	-0.19	-0.33	0.24	fb. & late noise red.
0.64	-0.09	1.17	0.53	0.14	noise red. & fb.

Figure 4.11: Model comparisons for group-level fits. A. residuals sum of squares for each model and cohort. Darker colours indicate lower values. B. Same as A for BIC. C-D. Estimatedfree parameter values β and γ for each model and cohort. Negative and positive values are indicated in blue and red, respectively. fb: feedback; noise red.: noise reduction; BIC: Bayesian information criterion.

residuals sum of squares (figure 4.11A) and lowest BIC (figure 4.11B) was the increased feedback with late noise reduction model—though the increased feedback model BIC was marginally lower for the large-reward cohort (ΔBIC= 3) and therefore was a similarly good fit. Finally, the β and γ parameters for the winning model both took negative values for all cohorts except the baselines cohort and the β parameter in the “large reward” cohort (figure 4.11C-D), in line with theoretical expectations.

Comparing group-level and invidividual-level model comparisons, we observe that the same model is consistently selected across all experimental cohorts besides the baselines cohort, corroborating the hypothesis that late noise reduction occurs alongside a global increase in feedback control in the presence of reward or punishment.

4.3.6 The effect of reward on end-point stiffness at the vicinity of the target

Next, we experimentally tested whether the reduction in noise observed in the late part of reward trials is associated with an increase in stiffness. For simplicity, we focused on the reward context only from this point. We recruited another set of participants (N=30) to reach towards a single target 20cm away from a central starting position in 0p and 50p conditions, and employed a well-established experimental approach to measure stiffness (Burdet et al., 2000; Selen et al., 2009). Specifically, during occasional "catch" trials (31% trials) a fixed-length (8mm) displacement was applied to the robotic manipulandum immediately as participants stopped within the target. These could be in 8 possible directions arrayed radially around the target (figure 4.12A). This displacement was transient, with a ramp-up, a plateau, and a ramp-down phase back to the original end-position. As the position was clamped during the plateau phase, velocity and acceleration were on average null, removing any influence of viscosity and inertia. Therefore, the amount of force required to maintain the displacement during plateau was linearly proportional to end-point stiffness of the arm (Perreault et al., 2002). The displacement profile of a participant is presented in figure 4.12B. Using a linear regression approach to fit the average recorded force during the plateau (grey area in figure 4.12B) against the displacement direction, we obtained the end-point stiffness matrices for all participants and all reward values. Stiffness matrices could then be visualised by plotting ellipses using the following equation:

$$\begin{bmatrix} x \\ y \end{bmatrix} = K \cdot \begin{bmatrix} \cos t \\ \sin t \end{bmatrix} \qquad 0 \leqslant t \leqslant 2\pi \qquad K = \begin{bmatrix} K_{xx} & K_{xy} \\ K_{xy} & K_{yy} \end{bmatrix} \tag{4.17}$$

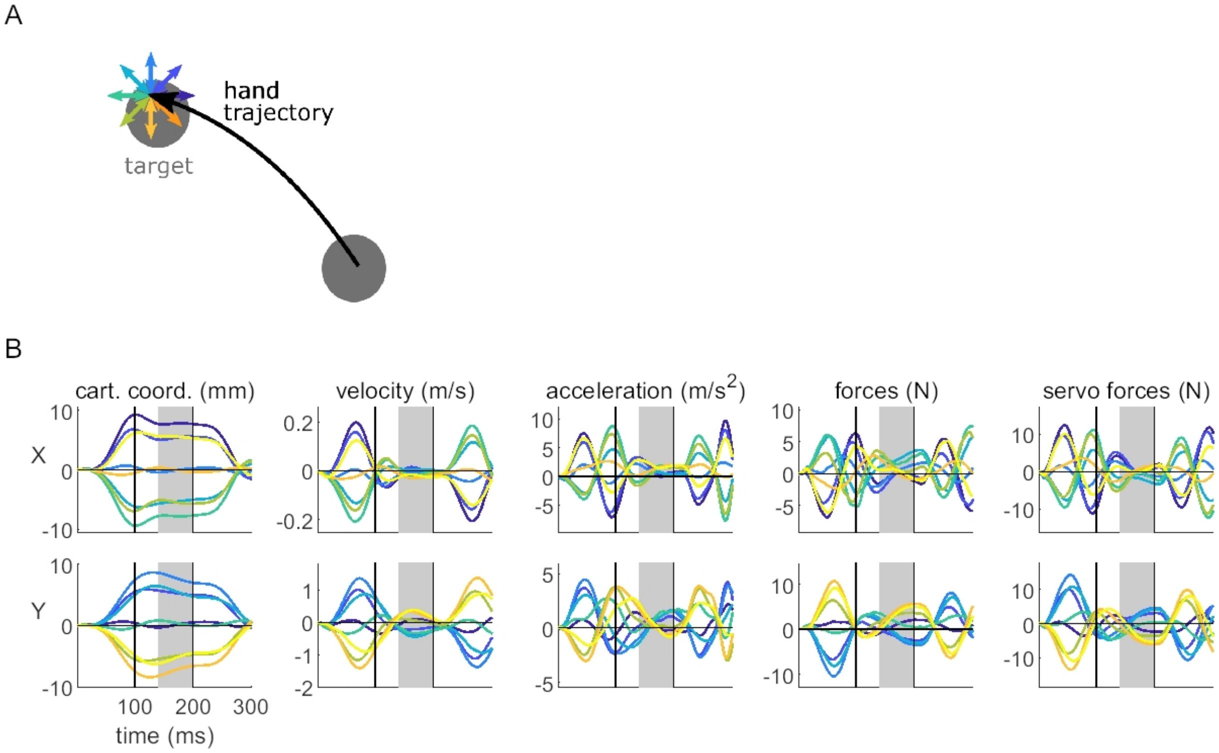


Figure 4.12: Displacement profile at the end of the reaching movement. A. Schematic of the displacement. At the end of the movement, when velocity decreased behind a threshold of 0.3 m/s, a displacement occasionally occurred in one of 8 possible directions. Each direction is represented by a colour. B. Average displacement profile over time for a sample participant. The upper and lower rows represent variables in the x and y dimension, respectively. The two vertical black solid lines demark the limit between the ramp-up and plateau, and plateau and ramp-down phase. Values for each variable were taken as the average over time during the 140-200ms window, where the displacement is clamped and most stable. This window is represented as a grey area in each plot.

Because arm stiffness is strongly dependent on arm configuration, stiffness ellipses are usually oriented, with a long axis indicating a direction of higher stiffness (figure 4.13). This orientation is influenced by several factors, including position in Cartesian space (Mussa-Ivaldi et al., 1985). If reward affects stiffness as we hypothesised, the possibility that this effect is dependent on a target location must therefore be considered. To account for this, two groups of participants (N=15 per group) reached for a target 45° to the right or the left of the starting position.

To quantify the global amount of stiffness, we compared the ellipse area across conditions (figure 4.13A-C). In line with our hypothesis, the area substantially increased in rewarded

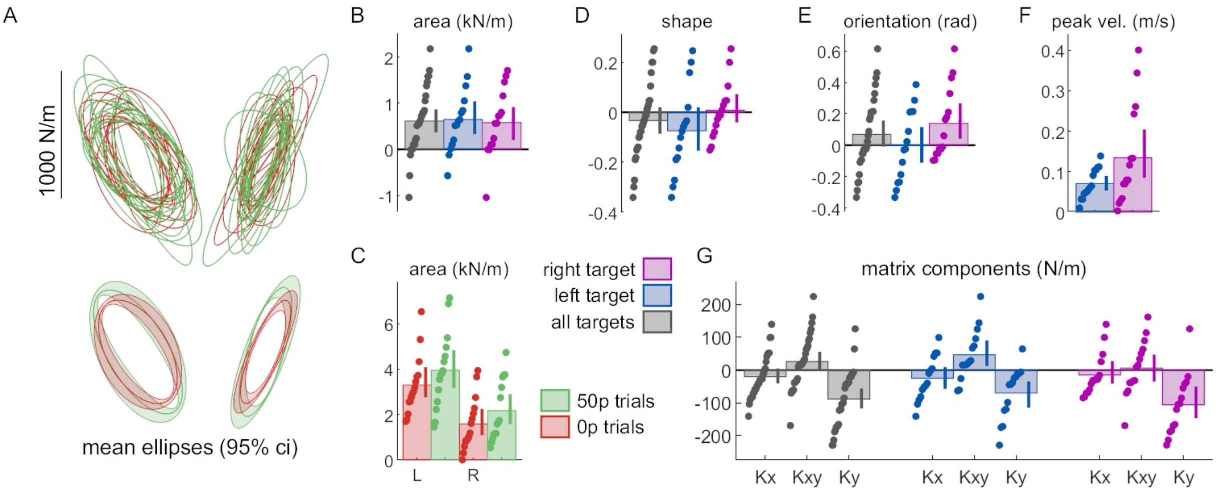


Figure 4.13: Reward increases stiffness at the end of movement. A. Individual (top) and mean (down) stiffness ellipses. Shaded areas around the ellipses represent bootstrapped 95% CIs. Right and left ellipses represent individual ellipses for the right and left target, respectively. B. Ellipses area normalised to 0p trials. Error bars represent bootstrapped 95% CIs. C. Non-normalised area values are also provided to illustrate the difference in absolute area as a function of target (L: left target, R: right target). D. Ellipse shapes normalised to 0p trials. Shapes are defined as the ratio of short to long diameter of the ellipse. E. Ellipse orientation normalised to 0p trials. Orientation is defined as the angle of the ellipse's long diameter. F. Peak velocity normalised to 0p trials. Peak velocity increased with reward. G. Stiffness matrix elements for 50p trials normalised to the stiffness matrix for 0p trials.

trials compared to non-rewarded trials (figure 4.13A,B). This effect of reward was very consistent across both target positions (figure 4.13B), even though absolute stiffness was globally higher for the left target (figure 4.13C). On the other hand, other ellipse characteristics, such as shape and orientation (figure 4.13D,E) showed less sensitivity to reward. However, since reward also increased average velocity (figure 4.13F), in line with our previous results, perhaps this increase in stiffness is a response to higher velocity rather than reward. To avoid this confound, we fitted a mixed-effect linear model, allowing for individual intercepts and target position intercept, where variance in area could be explained both by reward and velocity: $area \sim 1 + reward + peak\,velocity + (1|participant) + (1|target)$. As expected, reward—but not peak velocity—could explain the variance in ellipse area (peak velocity: $p = 0.31$; reward: $p < 0.001$; table 4.1), confirming that the presence of reward results in higher global stiffness at the end of the movement. In contrast, fitting a model with the same explanatory variables to the Ky component of the stiffness matrices, which showed the greatest sensitivity to reward compared to the other components (figure 4.13G) revealed that not only reward ($p < 0.001$, Bonferroni corrected) but also peak velocity (p=0.016, Bonferroni-corrected; table 4.2) explained the observed variance (model: $Ky \sim 1 + reward + peak\,velocity + (1|participant) + (1|target)$). In comparison, no significant effects were found to relate to the Kx component (reward: $p = 0.14$, peak velocity: $p = 1$, Bonferroni-corrected; $Kx \sim 1 + reward + peak\,velocity + (1|participant) + (1|target)$).

Model:
area $\sim$ 1 + velocity + reward + (1|target) + (1|participant)

Number of observations	60	*AIC*	1562.1
Fixed effects coefficients	3	*BIC*	1574.6
Random effects coefficients	32	*Log-Likelihood*	-775.03
Covariance parameters	3	*Deviance*	1550.1

Fixed effects coefficients (95% CIs):

variable	*estimate*	*SE*	*t-statistic*	*DF*	*p-value*	*lower CI*	*upper CI*
intercept	$1.58e^{+5}$	$1.09e^{+5}$	1.4411	57	0.15501	-61456	$3.77e^{+5}$
velocity	84461	83260	1.0144	57	0.31467	-82266	$2.51e^{+5}$
reward	52737	15180	3.4741	57	0.00099	22340	83134

Random effects covariance parameters (95% CIs):

variable	*levels*	*type*	*estimate*	*lower CI*	*upper CI*
target	2	std	89384	28576	279590
participant	30	std	1.2749	96198	$1.69e^{+5}$
error	60	residual std	48540	37688	62518

Table 4.1: Mixed-effect model for stiffness area at the vicinity of the target.

Because interactions with nested elements cannot be compared directly using a mixed-effect linear model (Schielzeth & Nakagawa, 2013; Zuur et al., 2010; Harrison et al., 2018), we employed a repeated-measure ANOVA to compare the interaction between reward and target on stiffness. No interaction between reward and target location were observed on area ($F(1, 28) = 0.069, p = 0.79$, partial $\eta^2 < 0.001$; figure 4.13A,C).

We conclude that end-point stiffness is sensitive to both reward and velocity. However, velocity-driven increase in stiffness is specific to the dimension that this velocity is directed toward, while reward-driven increase in stiffness is non-directional, at least in our task.

Model: **Ky ∼ 1 + velocity + reward + (1|target) + (1|participant)**							
Number of observations	60				*AIC*	731.43	
Fixed effects coefficients	3				*BIC*	743.99	
Random effects coefficients	32				*Log-Likelihood*	-359.71	
Covariance parameters	3				*Deviance*	719.43	
Fixed effects coefficients (95% CIs):							
variable	*estimate*	*SE*	*t-statistic*	*DF*	*p-value*	*lower CI*	*upper CI*
intercept	-178.28	80.817	-2.206	57	0.031432	-340.11	-16.447
velocity	-205.92	75.341	-2.7331	57	0.008341	-356.78	-55.049
reward	-66.893	16.903	-3.9575	57	0.000212	-100.74	-33.046
Random effects covariance parameters (95% CIs):							
variable	*levels*	*type*	*estimate*	*lower CI*	*upper CI*		
target	2	std	$8.60e^{-5}$	*NA*	*NA*		
participant	30	std	107.1	79.9	143.6		
error	60	residual std	58.18	45.16	74.94		

Table 4.2: Mixed-effect model for stiffness *Ky* component at the vicinity of the target.

This is likely because our task does not distinguish direction of error (*i.e.* error in the y dimension is not more punishing than in the x dimension) and so error must be reduced in all dimensions (Selen et al., 2009).

4.3.7 The effect of reward on end-point stiffness at the start of the movement

Finally, the time-time correlation maps also suggest that the increase in stiffness should only occur at the end of the reaching movement, since the early and middle parts show an opposite effect (decorrelation). Therefore, an increase in end-point stiffness should not be present immediately before the reach. To test this, participants (N=20) reached to 2 targets positioned 20cm away and 45° to the left and right of the starting position. On occasional catch trials (31% trials), a displacement akin to the previous experiment occurred in one of 8 possible directions at the time normally corresponding to target onset but after the reward information had been displayed. Unlike the previous experiment, reward and velocity had no impact on stiffness, either by area (reward: $p = 0.35$; peak velocity: $p = 0.75$, table 4.3) or by the matrix component Ky (reward: $p = 0.19$; peak velocity: $p = 0.45$, table 4.4), corroborating our interpretation of the correlation map (figure 4.15).

Model:
area ∼ 1 + velocity + reward + (1|participant)

Number of observations	40	*AIC*	1000.4
Fixed effects coefficients	3	*BIC*	1009.9
Random effects coefficients	20	*Log-Likelihood*	-495.22
Covariance parameters	2	*Deviance*	990.45

Fixed effects coefficients (95% CIs):

variable	*estimate*	*SE*	*t-statistic*	*DF*	*p-value*	*lower CI*	*upper CI*
intercept	176720	105090	1.6817	37	0.10106	-36206	389640
velocity	-34147	106840	-0.3196	37	0.75107	-250630	182330
reward	11547	12086	0.95537	37	0.34559	-12942	36036

Random effects covariance parameters (95% CIs):

variable	*levels*	*type*	*estimate*	*lower CI*	*upper CI*
participant	20	std	104260	75922	143160
error	*NA*	residual std	22268	16332	30360

Table 4.3: Mixed-effect model for stiffness area at the start of the movement.

Model:

$\textbf{Ky} \sim \textbf{1} + \textbf{velocity} + \textbf{reward} + (\textbf{1}|\textbf{participant})$

Number of observations	40	*AIC*	460.82
Fixed effects coefficients	3	*BIC*	469.27
Random effects coefficients	20	*Log-Likelihood*	-225.41
Covariance parameters	2	*Deviance*	450.82

Fixed effects coefficients (95% CIs):

variable	*estimate*	*SE*	*t-statistic*	*DF*	*p-value*	*lower CI*	*upper CI*
intercept	-421.01	134.26	-3.188	37	0.0029121	-700.04	-155.98
velocity	184.74	138.08	1.3379	37	0.18909	-95.041	464.53
reward	-12.34	16.319	-0.75617	37	0.45434	-45.406	20.726

Random effects covariance parameters (95% CIs):

variable	*levels*	*type*	*estimate*	*lower CI*	*upper CI*
participant	30	std	97.543	70.244	135.45
error	*NA*	residual std	32.425	23.767	44.237

Table 4.4: Mixed-effect model for stiffness *Ky* component at the start of the movement.

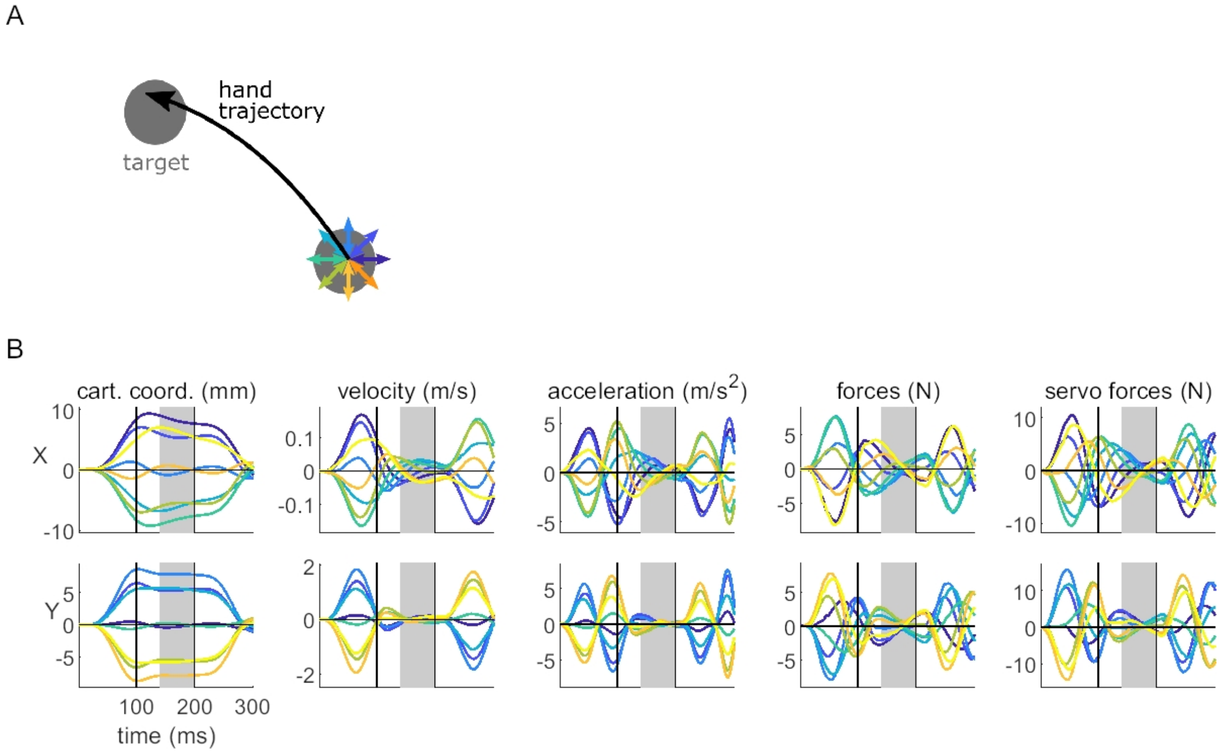


Figure 4.14: Displacement profile at the start of the reaching movement. A. Schematic of the displacement. At the start of the movement, a displacement occasionally occurred in one of 8 possible directions. Each direction is represented by a colour. B. Average displacement profile over time for a sample participant. The upper and lower rows represent variables in the x and y dimension, respectively. The two vertical black solid lines demark the limit between the ramp-up and plateau, and plateau and ramp-down phase. Values for each variable were taken as the average over time during the 140-200ms window, where the displacement is clamped and most stable. This window is represented as a grey area in each plot.

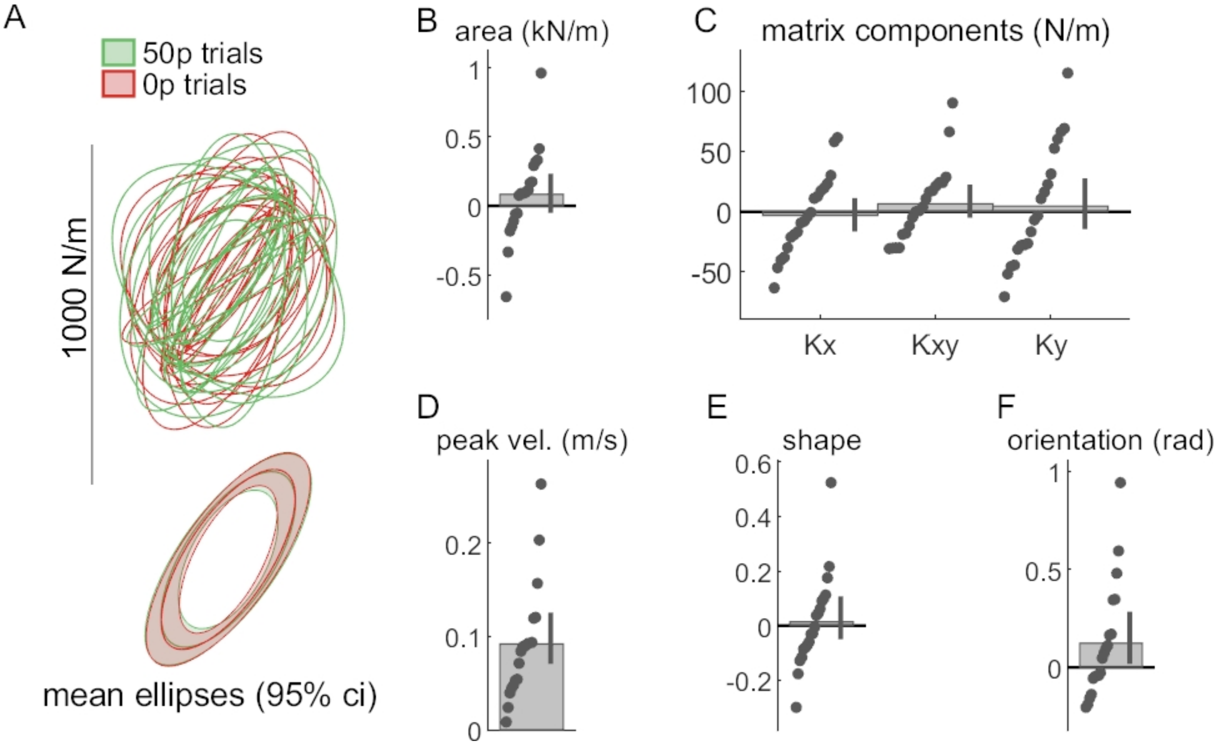


Figure 4.15: Reward does not alter stiffness at the start of movement. Individual (top) and mean (down) stiffness ellipses. Shaded areas around the ellipses represent bootstrapped 95% CIs. Right and left ellipses represent individual ellipses for the right and left target, respectively. B. Ellipses area normalised to 0p trials. Error bars represent bootstrapped 95% CIs. C. Stiffness matrix elements for 50p trials normalised to the stiffness matrix for 0p trials. D. Peak velocity normalised to 0p trials. E. Ellipse shapes normalised to 0p trials. Shapes are defined as the ratio of short to long diameter of the ellipse. F. Ellipse orientation normalised to 0p trials. Orientation is defined as the angle of the ellipse's long diameter.

4.4 Discussion

In this study, we showed that reward has the ability to simultaneously improve the selection and execution components of a reaching movement. Specifically, reward promoted the selection of the correct action in the presence of distractors, whilst also improving execution through increased speed and maintenance of accuracy. These results led to a shift in the speed-accuracy functions for both selection and execution. In addition, punishment had a similar impact on action selection and execution, though its impact was non-contingent for execution, in that it enhanced performance across all trials irrespective of feedback type. Computational analysis revealed that the effect of reward on execution involved a combination of increased feedback control and noise reduction, which we then showed was due to an increase in arm stiffness at the vicinity of the target—but not at the start of the movement. Overall, we confirm previous observations that feedback control increases with reward and offer a new error-managing mechanism that the control system employs under reward, in the form of arm stiffness regulation.

Our results add up to previous literature showing that reward increases execution speed in reaching (Chen, Holland & Galea, 2018; Pasquereau et al., 2007; Summerside et al., 2018) and saccades (Manohar et al., 2019, 2015; Takikawa et al., 2002). However, our results deviate from several reports in some respects. First, in a serial reaction time study, it was demonstrated that reward and punishment both reduce reaction times in humans (Wachter et al., 2009), while reaction times are not significantly altered by reward and punishment in our study. However, serial reaction time tasks strongly emphasise reaction times as a measure of learning independently of other variables, and interestingly, they show that punishment also led to a non-contingent effect on performance, while reward did not, similar to our results. A possible interpretation is that the motor system presents a similar bias to punishment to what is regularly reported in prospect theory and decision-making literature (Kahneman & Tversky, 1979; Chen et al., 2017a)—a phenomenon dubbed “loss aversion”. Next, radial accuracy has been shown to improve with reward, both in monkeys (Kojima & Soetedjo, 2017; Takikawa et al., 2002) and humans (Manohar et al., 2019, 2015), but these studies all focused on saccadic eye movements. In contrast, one reported case in a reaching task showed improvements in angular accuracy (Summerside et al., 2018). However, accuracy requirements in their no-reward condition were minimal, possibly allowing for larger improvements to be expressed compared to our task, potentially explaining why we do not observe similar improvement of radial or angular accuracy. Finally, while other studies have shown that speed-accuracy functions can shift with practice (Reis et al., 2009; Telgen et al., 2014), it is noteworthy that reward has a capacity to do so in what seems a nearly instantaneous time-scale, that is, from one trial to the next. Indeed, trials bearing different reward values were randomly intertwined in our study, meaning that this shift occurs within one trial. In contrast, the shift in speed-accuracy function observed with motor learning can take hours or even days to occur (Telgen et al., 2014).

4.4.1 Implications of increased stiffness with reward

While it is well established that stiffness has a beneficial effect on motor performance, our work provides the first set of evidence that this mechanism is employed in a rewarding context. Stiffness is likely regulated through a change in co-contraction of antagonist muscles, which is a simple but costly method to increase stiffness and enhance performance against noise (Gribble et al., 2003; Selen et al., 2009; Ueyama & Miyashita, 2013; Ueyama et al., 2011). The presence of reward may make such cost "worthy" of the associated metabolic expense (Todorov, 2004), as has been shown in saccades using an optimal control framework (Manohar et al., 2019, 2015) and in reaching in non-human primates (Ueyama & Miyashita, 2014). Nevertheless, the contribution of stiffness in reward-based performance has implications for current lines of research on clinical rehabilitation that focus on ameliorating rehabilitation procedures using reward (Goodman et al., 2014; Quattrocchi et al., 2017). While several studies report promising improvements, excessive stiffness may expose vulnerable clinical populations to increased risk of fatigue and even injury. Careful monitoring is therefore required to avoid this possibility.

4.4.2 Saccades and reaching movements differ in their use of stiffness control with reward

Contrary to our findings, previous work on saccade suggests that no effect of reward on stiffness can be observed (Manohar et al., 2019). Therefore, our results demonstrate that reaching movements differ from saccadic control, in that it employs an additional error-managing mechanism. Why do saccadic and limb control employ dissociable control approaches?

A first explanation may be the difference in motor command profile. Saccadic control displays a remarkably stereotyped temporal pattern of activity, in which the saccade is initiated by a transient burst of action potentials from the motoneurons innervating the extraocular muscles (Joshua & Lisberger, 2015; Robinson, 1964). Critically, this burst of activity always reaches an output frequency close to its maximum nearly instantaneously in an all-or-nothing fashion (Joshua & Lisberger, 2015; Robinson, 1964), with only marginal variation based on reward and saccade amplitude (Manohar et al., 2019; Reppert et al., 2015; Robinson, 1964; Xu-Wilson et al., 2009). In comparison, motor commands triggering reaching movements present a great diversity of temporal profiles depending on task requirements, and often do not reach maximum stimulation level. This difference between the two controllers may result in a difference in the temporal pattern of motor unit recruitment. According to the size principle (Llewellyn et al., 2010), low-force producing, high-sensitivity motor units are always recruited first during a movement. However, those motor units are also more noisy due to their higher sensitivity (Dideriksen et al., 2012). Since saccades always rely on an all-or-nothing input pattern, all motor units may be quickly recruited, including high-force, low-sensitivity motor neurons that are normally recruited last. This would drastically reduce the production of peripheral noise, thus making co-contraction unnecessary (Dideriksen et al., 2012). This is in line with previous work showing peripheral noise has a minimal contribution to overall error in eye movements (Van Gisbergen et al., 1981) compared to internally generated noise (Manohar et al., 2019). Interestingly, evidence of the opposite has been reported for reaching, suggesting that execution rather than planning noise is dominant in reaching errors (van Beers et al., 2004). These dissociable activation patterns of motor commands could potentially explain the differences in error-managing mechanisms between saccadic control and reaching.

A second possibility is that the muscles considered in saccade and reaching have different size and innervation density. Although eyes muscles are smaller, they are remarkably more innervated than most peripheral skeletal muscles (Floeter, 2010; Porter et al., 1995) such as arm muscles recruited for reaching, leading to a larger amount of motor units. Interestingly, it has been shown that motor noise arising at the muscle level scales negatively with the number of motor units in that muscle (Hamilton et al., 2004). This may lead to reduced levels of execution noise for eye movements compared to reaching movements, making stiffness regulation less necessary for saccades.

4.4.3 Increased feedback control and reward

It is less clear what kind of feedback control may play a role in reward-driven improvements. Feedback control encompasses several processes that have in common the tracking of deviation from a motor plan to correct for it, with varying amount of delay to allow for travelling from the peripheric sensory receptors to the brain. This includes the spinal stretch reflex (∼25ms delay; Weiler et al., 2019), transcortical feedback (∼50ms; Pruszynski et al., 2011) and visual feedback (∼170ms for fast involuntary visual feedback responses; Carroll et al., 2019). While spinal stretch reflex is extremely fast, it is difficult to assume an effect of reward or motivation occurring at the spinal level. On the other hand, transcortical feedback includes primary motor cortex processing (Pruszynski et al., 2011), a structure that shows sensitivity to reward (Bundt et al., 2016; Galaro et al., 2019; Thabit et al., 2011). Therefore, an exciting possibility for future research is that transcortical feedback gain is directly enhanced by the presence of reward. Indirect evidence suggests that this may be the case, as feedback control of matching timescales is sensitive to urgency in reaching (Crevecoeur et al., 2013). This suggests that transcortical feedback gains can also be pre-computed before movement initiation to meet task demands. Finally, recent work shows that reward can indeed modulate visual feedback control in reaching (Carroll et al., 2019) at timescales of 170-220ms after movement onset, much faster than usually considered for this type of feedback control (Carroll et al., 2019; Kasuga et al., 2015). Despite this remarkable speed, considering our typical movement times, this would imply that feedback control is increased only after about half of the movement. Therefore, a more conservative possibility is that both transcortical and visual feedback gains increase in the presence of reward, though the former remains to be proved empirically.

In saccades, it has been shown that the feedback controller that underlies reward-driven improvements is located further upstream, at the movement computation stage. Indeed, although saccadic control is ballistic and therefore feedforward, it has been shown that the cerebellum can provide some form of feedback to adjust the end part of a saccade trajectory based on errors in the forward model prediction (Chen-Harris et al., 2008). More recently, Manohar et al. (2019) demonstrate that it is this feedback loop that accounts for observed improvements in feedback control during saccades. Interestingly, evidence in humans show that cerebellar forward models do contribute to feedback control in reaching (Miall et al., 2007), and more recently, optogenetics manipulation in mice confirmed this suggestion (Becker & Person, 2019). Therefore, it is possible that reward also enhance this feedback loop, though this would only contribute to reducing noise arising at the higher, computational stage rather than at effector stage (Manohar et al., 2019).

4.4.4 Limitations of the model

The model we employ presents several assumptions and limitations. First, it reduces the movement to errors over time, because it only deals with the deviation from zero. This is similar to assuming that a perfect knowledge of the movement to be performed is already acquired, because deviations are only a function of the noise term. Furthermore, since the model is concerned with maintaining the system at a given value rather than "travelling" to a novel position, the expected bell-shaped profile of motor commands (Shadmehr & Krakauer, 2008; Todorov, 2004) is abstracted away, and thus the noise term is not signal-dependent (Todorov, 2005). These simplifications can be overlooked when considering model selection, because it is only concerned about a directional change from an arbitrary control model (*i.e.* increase versus decrease in time-time correlation). However, it may impede reliable parameter estimation because it remains an abstraction that excludes particular features such as two-dimensional reaches or signal-dependent noise—in line with the observed uncertainty of parameter estimations (4.10C-D). Future work using simulations based on a more complete model of the arm may provide further information regarding the evolution of saccadic and reaching profiles over time and allow better parameter estimation.

4.4.5 Neuromodulators mediating the reward-driven effects

A natural supposition is that dopamine mediates some, if not all of these effects. In the motor control literature, dopamine has been shown to mediate reward-based increases in vigour similar to our observations. For example, Parkinson's disease (PD) patients, who present reduced dopamine levels due to their pathology, display a tendency to execute slower movements (Manohar et al., 2015; Mazzoni et al., 2007). In addition, neural recordings in monkeys show that dopamine neuron activity can predict the vigour of an upcoming movement (da Silva et al., 2018). Regarding the selection process, PD patients do not exhibit an improvement with reward in saccades while age-matched healthy controls do (Manohar et al., 2015). Of note, however, this is mainly due to higher reaction times, as the error rate—that is, propensity to move towards a distractor target—remains globally similar to controls. Since our results show selection improvements through an increase in selection accuracy but not through a change in reaction times, this suggests that our observation of increased selection accuracy is not dopamine-driven. Another possibility is that visual processing of information, which is at the core of accurate target selection since targets are discriminated based on visual features, is enhanced at the occipital level though cholinergic modulation. There is pharmacological and electrophysiological evidence that cholinergic afference for the basal forebrain to visual cortices improve encoding of visual information in rat, by reducing between-cell correlation (reducing redundancy) while increasing within-cell reliability across trials (Goard & Dan, 2009; Pinto et al., 2013). Basal cholinergic activity is usually related to heightened attentional effort (Sarter et al., 2005), which monetary reward is known to induce (Hübner & Schlösser, 2010). This would suggest that reward has a global, cross-modality impact on the central nervous system that is mediated by multiple neuromodulators rather than by dopamine alone. Future work with PD patients or with pharmacological manipulation of cholinergic states may help answer some of these questions regarding the role of each neuromodulator.

4.4.6 Conclusion

In this study, we show that reward can improve the selection and execution components of reaching movement without interference. While we confirm previous suggestions that enhanced feedback control drives this improvement, we introduce a novel, peripheral rather than central mechanism by showing that global end-point stiffness is regulated by the monetary value of a given trial. Therefore, reward drives multiple error-reduction mechanisms which enable individuals to invigorate motor performance without compromising accuracy.

Chapter 5

NO EFFECT OF TRANSCRANIAL MAGNETIC STIMULATION OF THE VENTRO-MEDIAL PREFRONTAL CORTEX AND SUPPLEMENTARY MOTOR AREA ON REWARD-DRIVEN ENHANCEMENT OF MOTOR CONTROL

This chapter is based on the task introduced in chapter 4. The hypothesis was defined and the study was designed conjointly by P.J. Holland, J.M. Galea and me. I then implemented the task, ran the experiment, performed the TMS stimulations, analysed the data and designed the figures. The interpretation of results was done conjointly by P.J. Holland, J.M. Galea and me. J.M. Galea provided the fundings and materials. I wrote the chapter that is presented here.

5.1 Introduction

In eye movements, reward has a well-known ability to invigorate motor control, enhance accuracy, and promote accurate action selection in the face of potential distractors (Kojima & Soetedjo, 2017; Manohar et al., 2019; Sohn & Lee, 2006; Takikawa et al., 2002). In chapter 4, we extended these behavioural findings from eye movements to reaching movements. Specifically, we found that reward enhanced selection by increasing participant's propensity to move towards the correct target in the presence of a distractor target, while reaction times were not impeded. Execution of reaching movements also showed a steep increase in peak velocity (vigour) with reward, while radial accuracy was maintained. While these effects are now behaviourally well-characterised and replicated, we now ask what neural substrates implement these phenomena. To this end, we aim at disrupting activity of specific cortical regions to underline a causal relationship with behaviour, using continuous theta-burst TMS (Y.-Z. Huang et al., 2005; Zenon et al., 2015).

During a sensorimotor task, a stream of information contributes to the generation of movement, coming from visual and proprio-tactile sensory afferents to high-level associative areas mainly in prefrontal regions, forming into a motor plan in the the supplementary motor area (SMA) and pre-motor area, to finally produce a motor command which travels down from the primary motor cortex (M1) to the spinal cord and to the effector muscles (Castiello, 2005; Hikosaka et al., 2002; Shadmehr & Krakauer, 2008). At which point of the sensory-prefrontal-premotor-motor loop does reward influence this processing stream, leading to behaviourally enhanced performance?

Sensitivity of attentional processes to reward availability is a well-known phenomenon (Sarter et al., 2006). For instance, reward-driven selection improvements similar to our observations have been reported in the flanker task (Hübner & Schlösser, 2010), a seminal paradigm for studying attentional capacity. The authors argued that information encoding may be enhanced with reward, drastically improving evidence accumulation and thus action selection. Such a mechanism has been showed to occur in visual cortices through cholinergic modulation in rats (Goard & Dan, 2009; Pinto et al., 2013), and imaging studies show that occipital regions exhibit the most sensitivity to reward in attentional tasks in human (Anderson, 2016; Tosoni et al., 2013). Thus, it may be that reward-driven selection improvements are due to early enhancement of visual sensory processing in the sensorimotor loop, a possibility that has been considered in a saccade study (Manohar et al., 2015). However, in that study, the authors also found that PD patients did not exhibit the increase in selection accuracy with reward seen in healthy aged-matched controls, suggesting that though acetylcholine may play a role in enhancement of selection accuracy, a role for dopamine should be considered as well. In line with this argument, a large number of imaging studies have demonstrated the involvement of the posterior and anterior cingulate cortices, and ventro-medial prefrontal cortex (vmPFC) in reward processing (Blair et al., 2013; Daw et al., 2005, 2006; Graybiel, 2008; Klein-Flugge et al., 2016), regions that are heavily dependent on dopamine innervation (Arnsten, 1998) and also involved in the sensorimotor loop (Hikosaka et al., 2002). Furthermore, magnetic resonance imaging (MRI) evidence shows the involvement of vmPFC in processing value of different stimuli during a decision-making task involving motor effort (Klein-Flugge et al., 2016). Thus, TMS on prefrontal or occipital areas should be considered. However, since occipital areas are not only involved in reward processing but also a large array of basic visual functions, TMS in these regions could potentially disrupt basic motor performance, and thus expose to unnecessary confounds. Therefore, we will focus on prefrontal manipulations in this study.

Regarding execution, are improvements due to enhanced encoding of visual information, thereby allowing more vigorous movements at no accuracy cost? Though this is a possibility, this would not explain the reward-driven increase in feedback control (Carroll et al., 2019; Manohar et al., 2019) and end-point stiffness we observe. Rather, reward could directly modulate M1, as its activity has been shown to be highly sensitive to reward (Bundt et al., 2016; Kapogiannis et al., 2008; Mawase et al., 2017, 2016; Ramkumar et al., 2016; Thabit et al., 2011; Zhao et al., 2018; Galaro et al., 2019), shaping performance at the very end of the processing steam. Another reasonable hypothesis is that reward information is integrated earlier on, with M1 being merely the final recipient. Several prefrontal regions upstream of M1 are involved in action planning, including the supplementary motor area (SMA), a region also showing strong sensitivity to reward (Klein-Flugge et al., 2016; Stanford et al., 2013; Zenon et al., 2015). In PD patients, who express apathy symptoms sometimes interpreted as a lack of vigour, also show altered SMA activity (Hendrix et al., 2018; Rascol et al., 1994). In recent work, it was argued that SMA encodes sensitivity to effort (Klein-Flugge et al., 2016), which is hypothesised to drive the change in vigour during motor control (Manohar et al., 2015; Mazzoni et al., 2007).

While TMS stimulation over M1 would not answer whether reward is integrated in M1 or earlier on, SMA stimulation could provide more conclusive evidence. If an effect on reward-driven enhancement of execution performance is seen, this would confirm that reward information is indeed integrated earlier than might be expected for reaching movements (Mawase et al., 2017, 2016; Thabit et al., 2011). Regarding the selection component, disrupting the anterior cingulate cortex or the vmPFC should alter reward-driven enhancements in selection. However, here, as the anterior cingulate cortex cannot be stimulated using TMS due to its deep location, we tested our hypothesis using the vmPFC.

5.2 Methods

5.2.1 Participants

26 of 34 screened participants (see section 5.2.4 for screening details) were selected based on their performance on the reaching task. Of those 26 selected participants, one was excluded due to medical reasons, and two participants retracted after the second session.

Therefore, 23 participants (median age: 22, range: 18-39, 15 female) took part in the whole experiment. All participants were right-handed, free of epilepsy, familial history of epilepsy, motor, psychological or neurological conditions, or any medical condition forbidding the use of TMS or MRI, in line with the rules of the University of Birmingham Ethics Committee. The study was approved by and done in accordance with the University of Birmingham Ethics Committee under the project code ERN_09-528P regarding the behavioural aspect of the experiments, and project code ERN_17-1541P regarding the TMS.

5.2.2 Task design

The behavioural task was identical to the first experiment of chapter 4, except that only 0p and 50p trials were used. During the screening session, participants first practiced the task in a 48 trials-long training block with a 25p trial value. They then performed a 16 trials-long, distractor-free baseline block with 0p and 50p trials and were informed that their score now added up toward their monetary gain. Finally, they experienced a 224 trials-long main block that included 96 (42.9%) distractor-containing trials randomly interspaced. For the three TMS sessions, participants repeated the same task, with the exception of the training block which was removed.

5.2.3 Procedure

The experiment took place over five sessions each at least five days apart from the previous one. The first session was a screening session, in which participants were selected based on their performance during the behavioural task. In the second session, a structural MRI scan of each participant's brain was acquired, and used for the third to fifth session, during which participants performed the behavioural task after receiving either sham, SMA or vmPFC theta-burst TMS (figure 5.1A). The order of stimulation was pseudo-randomly

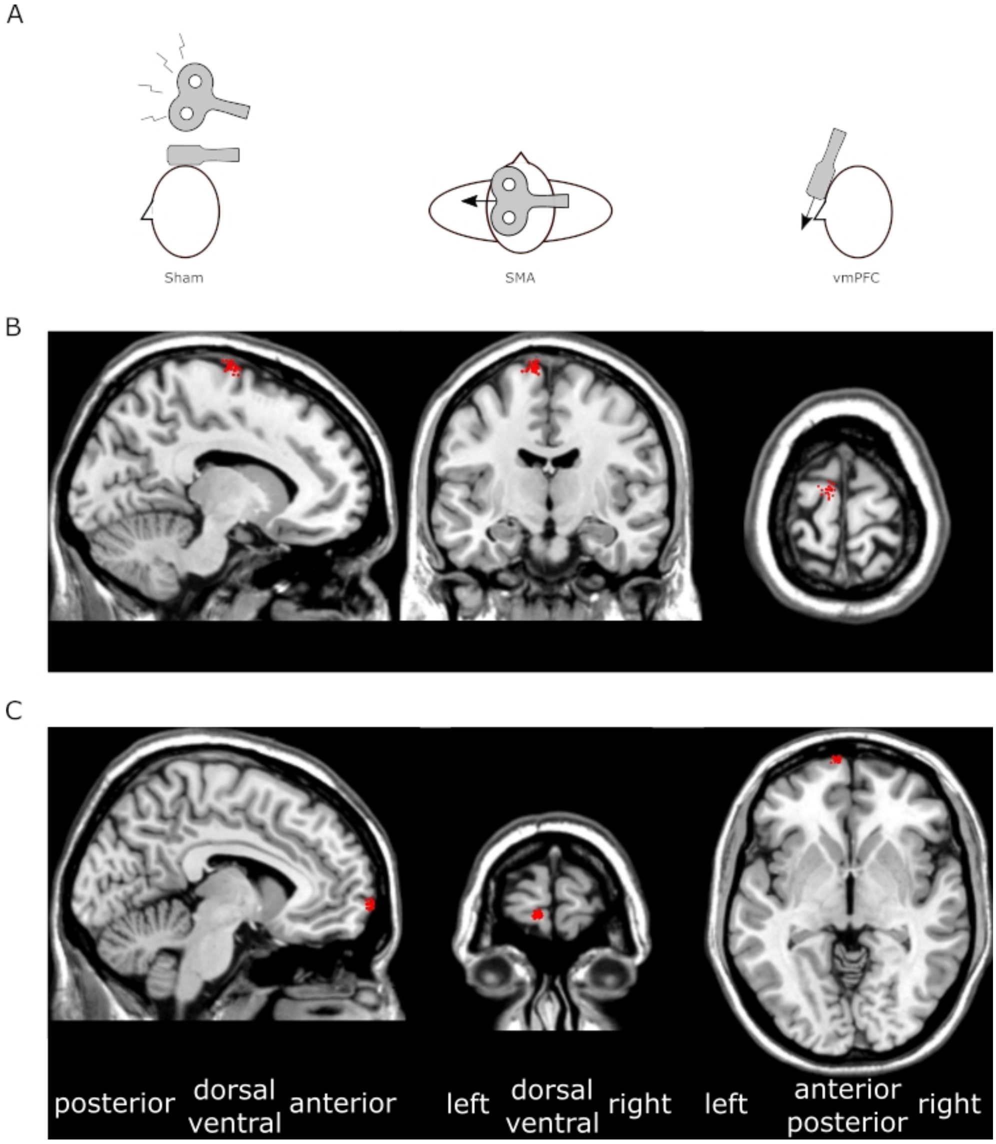


Figure 5.1: TMS procedure. A. position of the TMS coil(s) relative to the head in each of the 3 conditions. The black arrows represent the current orientation. B. Sagittal, coronal and axial planes of an MNI-normalised brain scan (*ch2.nii.gz* in MRIcron). The red dots indicate each participant's SMA stimulation sites. C. vmPFC stimulation sites.

counterbalanced across participants. Before every session, participant's health condition was assessed in accordance to the guidelines of the Ethics Committee of the University of Birmingham (UK).

5.2.4 Screening session

Because this study is interested in manipulating a previously isolated effect, we screened participants based on the presence of that effect to increase statistical power. For their first session, participants were first screened for medical or psychological conditions that could exclude them from the study. They were then introduced to the TMS technique by reading a leaflet, and they could ask any questions they wanted to the experimenter. Next, they were exposed to theta-burst stimulation on their forearm to get acquainted with the stimulation sensation. Their active motor threshold (AMT) was then determined by finding the single-pulse TMS intensity that resulted in the smallest possible twitch on their right-hand index finger five times out of ten. During the single-pulse TMS procedure, the coil was oriented at $-45°$ from the midline. Finally, participants performed the behavioural task. Using the resulting behavioural data, participants were then screened for an effect of reward on execution and selection accuracy. Specifically, participants were expected to show an increase in peak velocity and selection accuracy (*i.e.* increased propensity to ignore a distractor target) in rewarded trials compared to non-rewarded trials. Participants who did not show both of these effects or showed an overly weak effect were excluded. Of note, no participant showed an effect opposite to the effect of interest.

5.2.5 TMS procedure

Using a 3-T Philips scanner, high-resolution T1-weighted images were acquired for each participant (1x1x1mm voxel size, 175 slices in sagittal orientation). The image was then normalised to an MNI template using an affine (12 parameter) transformation (Jenkinson et al., 2002; Jenkinson & Smith, 2001) with the software Statistical Parametric Mapping 12 (SPM12). Regions of interest were then marked using MRIcron (Rorden & Brett, 2000). The MNI coordinates used were $x = -8/y = -9/z = 77$ for the SMA and $-7/71/-4$ for the vmPFC (figure 5.1B, C). More specifically, the SMA target region was the posterior part of the superior frontal gyrus, or the most prominent posterior part of Brodmann area 6 (Arai et al., 2012; Zenon et al., 2015); the vmPFC target region was the most anterior part of medial orbitofrontal gyrus, or Brodmann area 10 near the limit with Brodmann 11 (Blair et al., 2013; Lev-Ran et al., 2012). These positions were all in the left hemisphere (Arai et al., 2012; Lev-Ran et al., 2012) since all our participants were right-handed. The marked scans were then transformed back into their original shape using each participant's inverse transform with SPM12, and the position of each mark was manually inspected and adjusted to the closest location minimising distance between the target position and the scalp (Galea et al., 2010; Y.-Z. Huang et al., 2005), giving subject-specific target locations.

TMS was applied using the continuous theta-burst stimulation technique, with one cycle lasting 40s, at 80% AMT or 48% intensity, whichever the lowest. A total of 200 burst trains were applied at a frequency of 5Hz, with 3 pulses per burst and a pulse frequency of 50Hz—giving a total amount of 600 pulses. These parameters were all based on (Galea et al., 2010; Y.-Z. Huang et al., 2005). During all TMS sessions (including the sham session), participants were asked if they felt fine immediately after the stimulation was performed, and upon confirmation, were asked to move from the stimulation chair to the KINARM chair on which they could perform the behavioural task. This represented a distance of approximately two meters.

5.2.6 Data analysis

The pre-registered *a priori* hypotheses, TMS procedure and planned analyses are all available online on the Open Science Framework website, alongside the experimental dataset used and analysis scripts (`https://osf.io/tnkrj/`). Analyses were performed using custom Matlab scripts (Matworks, Natick, MA). All trial-by-trial analyses were identical to that of chapter 4 and are repeated here for convenience.

Trials were manually classified as distracted or non-distracted. Trials that did not include a distractor target—*i.e.* no-distractor trials—were all considered non-distracted. Distracted trials were defined as trials where a distractor target was displayed, and participants initiated their movement (*i.e.* exited the starting position) toward the distractor instead of the correct target (see figure 4.2). If participants readjusted their reach "mid-flight" to the correct target or initiated their movement to the right target and readjusted their reach to the distractor, this was still considered a distracted trial.

Reaction times were measured as the time between the correct target onset and when the participant's distance from the centre of the starting position exceeded 2cm. In trials that were marked as "distracted" (*i.e.* participant initially went to the distractor target), the distractor target onset was used. In distractor-containing trials, the second, correct target did not require any selection process to be made, since the appearance of the distractor target informed participants that the next target would be the right one. For this reason, reaction times were biased toward a faster range in non-distracted trials. Consequently, mean reaction times were obtained by including only no-distractor trials, and distracted trials (figure4.2). For every other summary variable, we included all trials that were not distracted trials, that is, we included non-distracted trials and no-distractor trials (figure4.2).

Trials with reaction times higher than 1000ms or less than 200ms, and non-distracted trials with radial errors higher than 6cm or angular errors higher than 20° were removed. Overall, this accounted for 0.49% of all trials. Speed-accuracy functions were obtained for each participant by binning data in the x-dimension into 50 quantiles and averaging all y-dimension values in a x-dimension sliding window of a 30-centile width (Manohar et al., 2015). Then, each individual speed-accuracy function was averaged by quantile across participants in both the x and y dimension.

Group statistics were performed using a 2x3 repeated-measure ANOVA, with reward value (0p versus 50p) as the first factor, and TMS group (sham, SMA, vmPFC) as the second factor. Effect sizes are all reported as partial η^2. Because main effects were only detected in the first factor (0p-50p) that presented 2 levels, no-post-hoc analyses were necessary for this chapter. For all plotted variables, bootstrapped 95% CIs of the mean were obtained using 10,000 permutations.

5.3 Results

5.3.1 Effect of reward on reaching performance

Similar to chapter 4, reward improved the selection and execution components of reaching movements (figure 5.2). Specifically, reward led to faster reaction times ($F(1,22) = 8.18, p = 0.009$, partial $\eta^2 = 0.37$; figure 5.2A), whilst also improving selection accuracy ($F(1,22) = 16.7, p < 0.001$, partial $\eta^2 = 0.76$; figure 5.2B), clearly demonstrating that the selection component benefited from the presence of reward. Regarding execution, peak velocity increased with reward ($F(1,22) = 42.4, p < 0.001$, partial $\eta^2 = 1.93$; figure 5.2C) whilst movement time decreased ($F(1,22) = 24.0, p < 0.001$, partial $\eta^2 = 1.09$; figure 5.2D). In addition, radial error ($F(1,22) = 2.88, p = 0.10$, partial $\eta^2 = 0.13$; figure 5.2E) and angular error ($F(1,22) = 2.98, p = 0.10$, partial $\eta^2 = 0.14$; figure 5.2F) were similar across rewarded and non-rewarded trials.

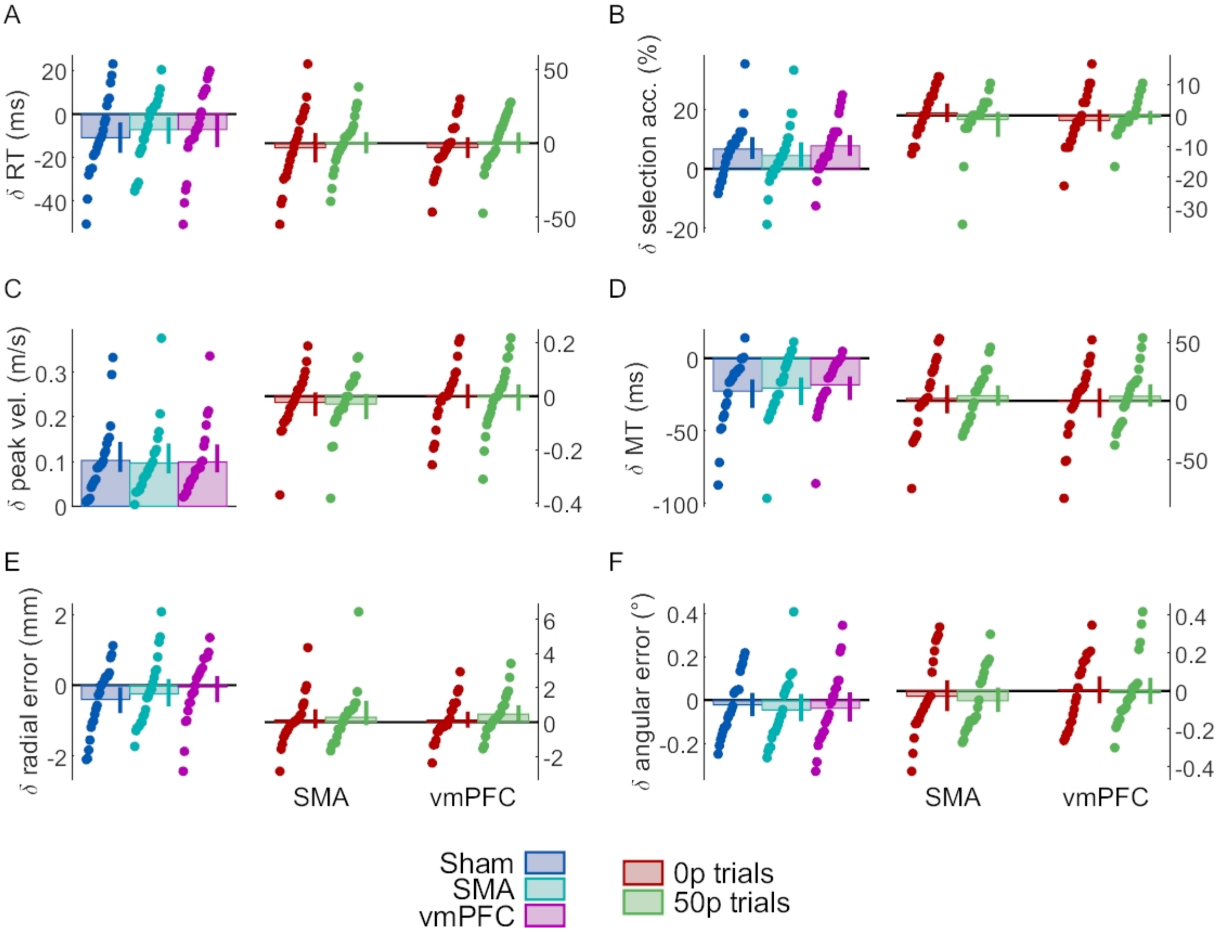


Figure 5.2: Effect of reward and TMS on different behavioural variables. A. Reaction times. On the left, 50p trials performance for each TMS group are normalised to 0p trials (*i.e.* reward-normalised), and on the right 0p and 50p trials for each TMS group are normalised to sham performance (*i.e.* sham-normalised). Dots represent individual values for each group and the error bars indicate bootstrapped 95% CIs of the mean. B-F. Other variables in the same format as panel A.

5.3.2 Effect of TMS manipulation on reward-driven behaviours

While we expected to observe an effect of TMS on the reward-driven effects, we observed no main or interaction effects for TMS: reaction times (TMS: $F(2, 44) = 0.05, p = 0.95$, partial $\eta^2 = 0.002$; interaction: $F(2, 44) = 0.65, p = 0.53$, partial $\eta^2 = 0.03$; figure 5.2A), selection accuracy (main effect of TMS: $F(2, 44) = 0.40, p = 0.70$, partial $\eta^2 = 0.02$; interaction: $F(2, 44) = 1.12, p = 0.33$, partial $\eta^2 = 0.05$; figure 5.2B), peak velocity (TMS: $F(2, 44) = 0.85, p = 0.43$, partial $\eta^2 = 0.04$; interaction: $F(2, 44) = 0.19, p = 0.83$, partial $\eta^2 = 0.008$; figure 5.2C), movement times (TMS: $F(2, 44) = 0.21, p = 0.81$, partial

$\eta^2 = 0.009$; interaction: $F(2,44) = 0.78, p = 0.46$, partial $\eta^2 = 0.03$; figure 5.2D), radial (TMS: $F(2,44) = 0.79, p = 0.46$, partial $\eta^2 = 0.04$; interaction: $F(2,44) = 1.08, p = 0.35$, partial $\eta^2 = 0.05$; figure 5.2E) and angular error (main effect of TMS: $F(2,44) = 1.18, p = 0.32$, partial $\eta^2 = 0.05$; interaction: $F(2,44) = 0.16, p = 0.86$, partial $\eta^2 = 0.007$; figure 5.2F). This indicates that continuous theta-burst TMS over vmPFC or SMA had no effect on behaviour.

5.3.3 Effect of reward and TMS on speed-accuracy functions

Next, we assessed the speed-accuracy functions of the selection and execution components in all TMS conditions. As can be seen in figure 5.3, we can reliably and consistently see a shift in the speed-accuracy functions of both these components with reward, in line with previous results (figure 5.3A-F). However, the execution speed-accuracy function in the SMA TMS group does not exhibit a normal profile at baseline (0p trials; figure 5.3E). Instead, radial error appears to be maintained across the range of peak velocities displayed. However, this profile did not extend to rewarded trials. Because this behaviour at baseline is surprising, we examined individual speed-accuracy profiles for this condition to ensure this was not driven by outliers. We can observe from figure 5.4 that indeed, two participants displayed more accurate performance at high speeds for 0p trials in the SMA TMS condition (middle panel), compared to the majority of participants. However, overall, there were also more participants who exhibited more accurate performance at higher speeds in this condition than in comparable conditions, such as 0p trials in the sham condition (figure 5.4, left panel) or the 50p trials in the SMA TMS condition (right panel). Therefore, while no clear speed-accuracy trade-off was observed for the 0p trials in the SMA TMS condition, it cannot be conclusively stated that this was driven by outliers.

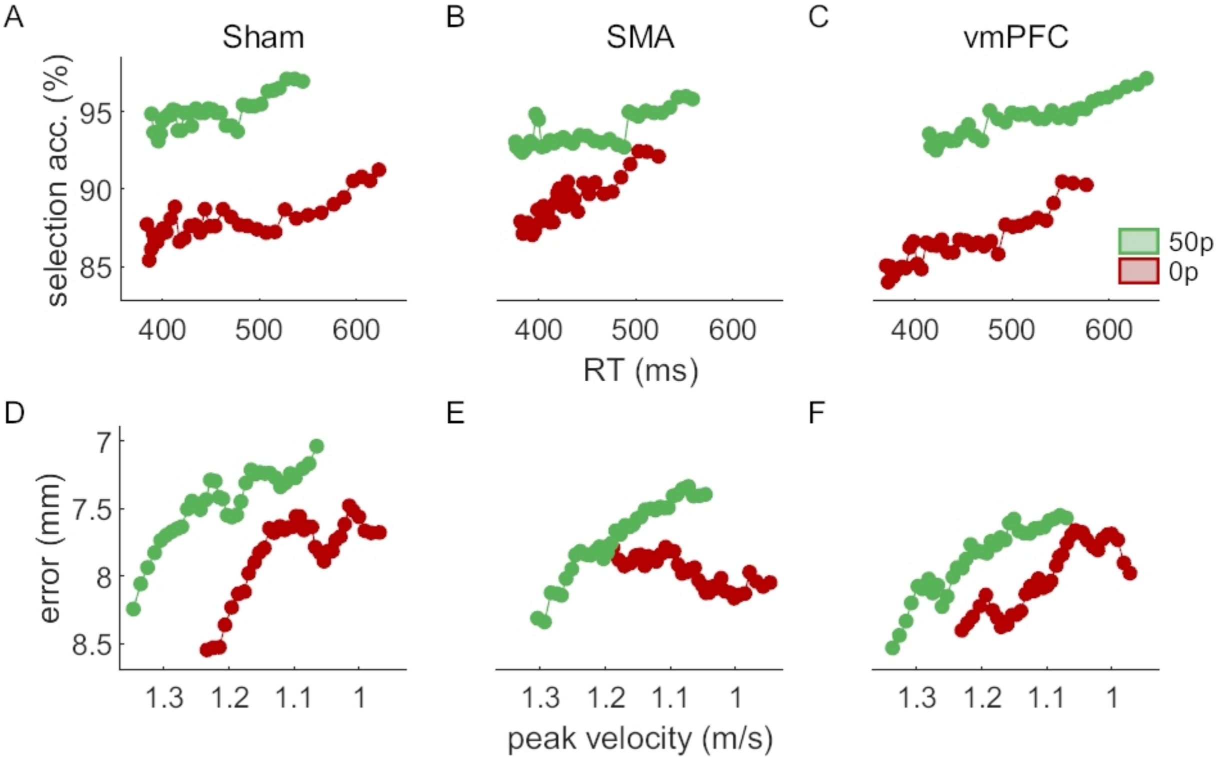


Figure 5.3: Speed-accuracy functions for each reward and TMS condition The selection (A-C) and execution (D-F) speed-accuracy functions are the top three and bottom three panels, respectively. The functions are obtained by sliding a 30% centile-wide window over 50 quantile-based bins and averaging each bin across participant. For the selection panels, the count of non-distracted trials and distracted trials for each bin was obtained, and the ratio (100*non-distracted/total) calculated afterwards. Note that the axes of the execution functions are reversed so that high speed and low accuracy are on the bottom-left corner like for the selection functions.

5.4 Discussion

In this study, we aimed at perturbating neural activity of the vmPFC and SMA using theta-burst TMS in an attempt to modulate previously characterised reward-driven effects on selection and execution performance in a reaching task. While the effects relating to reward from chapter 4 were reliably reproduced within-participant and across a series of four sessions held in different days, theta-burst TMS stimulation on either of the two target regions did not result in any alteration of these effects.

The replication of previous findings regarding the effect of reward on a reaching task across weekly sessions and on the same individuals strengthens the conclusions from chapter 4.

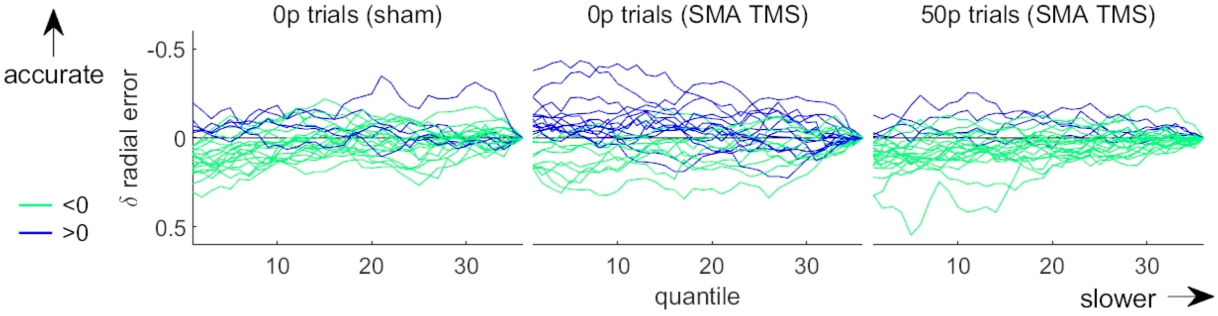


Figure 5.4: Individual speed-accuracy functions for the no-reward condition of the SMA TMS group (middle) and for two control groups (right and left). The functions are obtained by sliding a 30% centile window over 50 quantile-based bins. Each individual profile is normalised to its end value. Profiles exhibiting an increase and a decrease in accuracy with slower movements are plotted in light green and blue, respectively.

While it could be argued that this is natural considering the pre-selection of participants, it was not granted that an effect found on one day for a given participant could replicate consistently in a subsequent session held on another day, and this is still a reassuring outcome in the context of the widespread replication crisis currently being debated in psychology and neuroscience (Baker, 2016; Open Science Collaboration, 2015). Nevertheless, one divergent result is the observed reduction in reaction times with reward in this chapter, as no significant effect had been observed in chapter 4. However, a similar trend that failed to reach significance had been observed. Here, pre-selecting participants may have allowed that trend to reach a significance threshold. This suggests that there is an effect of reward on reaction times, although likely small in size.

The absence of significant effects of TMS on the reward-driven effects is more problematic. As the aphorism goes "*the absence of evidence is not evidence of absence*" and so drawing conclusions on the sole basis of non-significant results is a well-established fallacy (Altman & Bland, 1995). Therefore, the rest of this discussion is merely speculative rather than conclusive, although it can provide additional information to back up previously reported evidence.

First, the absence of an effect of vmPFC stimulation could suggest that other regions may influence the selection component of motor control. As mentioned previously, early sensory areas such as visual cortices are possible candidates (Anderson, 2016; Goard & Dan, 2009; Pinto et al., 2013; Tosoni et al., 2013). However, prefrontal regions show a very complex hierarchical organisation for reward information processing (Hunt & Hayden, 2017), and other possibilities should not be overlooked. It could be for instance that other well-known reward-processing centres located in the prefrontal areas are involved in processing the selection aspects of motor control, such as the cingulate cortex (Blair et al., 2013; Klein-Flugge et al., 2016; Tosoni et al., 2013), which is unfortunately not a possible target for TMS stimulation due to its deep location. Another possibility is that vmPFC is indeed involved in the selection process, but that the processing network allows for some compensatory activity, meaning that perturbing vmPFC activity does not affect the network capacity as a whole. Finally, it could be that vmPFC is involved in selection but TMS is not as effective in perturbing neural activity in vmPFC as in other regions. To our knowledge, only one study reports a significant effect of repetitive TMS on vmPFC (Lev-Ran et al., 2012), suggesting that perturbation of neural activity with this technique remains possible—though it cannot be ascertained whether our specific stimulation protocol or task design can successfully do so. While that study stimulated participants every 15 minutes, the experiment presented here lasted about 15 minutes as well, suggesting that an effect would have sustained long enough over time. Overall, it is not clear whether the reliably observed effects triggered by M1 theta-burst TMS can generalise to vmPFC stimulation.

The situation is less ambiguous regarding the absence of an effect of TMS stimulation on SMA. First, there are numerous studies showing theta-burst TMS influences SMA activity (Zenon et al., 2015; Arai et al., 2011, 2012; Matsunaga et al., 2005; Shirota et al., 2012), some of them showing that stimulation can also modulate downstream regions such as M1 (Arai et al., 2011, 2012; Matsunaga et al., 2005; Shirota et al., 2012). This last point indicates that any TMS effect was strong enough to lead to consequences even in regions that were not directly stimulated, and overall, the literature demonstrates that theta-burst TMS does generalise to SMA. However, due to the "drawer effect" bias (Open

Science Collaboration, 2015), it is difficult to estimate to what extent this manipulation can reproducibly perturb neural processing of SMA. Nevertheless, considering the large set of available studies showing a significant effect of TMS , it is more plausible that other regions implement reward-driven effects on execution than to assume that TMS is ineffective in manipulating SMA activity. Mainly the pre-motor area and M1 represent potential alternative candidates. However, while the premotor area is central to movement planning, it has shown sensitivity to motivation but not to reward compared to SMA (Ramkumar et al., 2016; Roesch & Olson, 2003, 2004). On the other hand, a large literature demonstrates effects of reward on various aspects of M1 processing (Bundt et al., 2016; Kapogiannis et al., 2008; Mawase et al., 2017, 2016; Ramkumar et al., 2016; Thabit et al., 2011; Galaro et al., 2019), making it a more suitable candidate for mediating reward-driven effects observed in our study. We show in chapter 4 that some execution improvements may be due to an increase in feedback control, likely transcortical (Omrani et al., 2016; Pruszynski et al., 2011) and visuomotor feedback (Carroll et al., 2019). Interestingly, transcortical feedback relies on M1 modulation (Pruszynski et al., 2011), in line with the possibility that M1 supports execution improvements.

Overall, this study shows that the reward-driven effects on reaching are robust and replicable across multiple sessions for a given participant. However, TMS on the vmPFC and SMA was ineffective in manipulating these effects. While it is difficult to interpret this absence of TMS effects, we outline possible explanations for this. Notably, the absence of effect following SMA TMS further bolsters the possibility that reward impacts motor execution at a late stage of the sensorimotor loop, likely at the level of M1.

Chapter 6

GENERAL DISCUSSION

6.1 Summary of results

In chapter 2, we consider the relationship between explicit control and reinforcement learning during a visuomotor rotation task. While we qualitatively reproduce results from previously reported results, we show that reward-induced effects can in fact be explained by explicit control, leading to a reconsideration of its role in reported reinforcement-based effects in motor learning.

In chapter 3, we ask what is the source of the significant variability that is regularly observed in reward-based motor learning and consider different working memory and genetic candidates. Spatial and mental rotation working memory explained most of the variability in the dataset, and against our expectation, individual dopamine-related genetic variability had no explanative power. While the working memory effect was in line with our results from chapter 2 showing a preponderant role of explicit control in reinforcement-based learning, the absence of any genetic effect refuted the possibility of a strong dopaminergic influence in this process.

In chapter 4, we turn to the question of motor control and how well-established reward-induced enhancements in motor control may be implemented mechanistically. We confirm previous observations that feedback control increases with reward and offer a new error-managing mechanism that the control system employs under reward, in the form of arm stiffness regulation.

Chapter 5 is concerned with the neural correlates of reward-based improvements in motor control. In this study, theta-burst repetitive TMS was performed on the vmPFC and SMA with the hope that they would disrupt reward-based effects on action selection and action execution compared to a sham control condition, respectively. However, no effect was observed in either condition.

6.2 Impact of this work: reappraising the role of explicit control in the literature

The introduction of reinforcement in the motor learning literature came together with the assumption that this was an implicit process in nature (Haith & Krakauer, 2013), that is, participants were unaware of and had no volitional control over its involvement. This assumption has at least two sources. First, the decision-making literature conceptualized parts of its reinforcement processes—namely, model-free reinforcement—as habit-like (Balleine & O'Doherty, 2010; Otto, Raio et al., 2013). When imported to the field of motor control, this led to an association with automaticity because skills exhibit automaticity. This association was on occasions more harmful than fruitful (Stanley & Krakauer, 2013), since habits are often considered implicit, while automaticity is not, as argued by Douskos (2017). To illustrate this difference, let us imagine someone performing a skill such as playing guitar. While performing, lack of focus will naturally lead to poorer performance. In contrast, habits exhibit the opposite relationship, that is, lack of focus will lead to more habitual slips while increased focus reduces habitual behaviour (Douskos, 2017). Second, the motor control field has long carried a view that motor learning is fundamentally an implicit compound of human memory, giving echo to this association. This stems from the classic work on H.M., a patient with retrograde amnesia who forgot overnight any episodic memory he formed on the day, but could still learn new motor skills with the same ease as healthy patients (Squire, 2004). This bias also led to the view that cerebellar adaptation was fundamentally an implicit mechanism, when in fact it is now accepted as a bi-component mechanism including both implicit and explicit contributions (Bond & Taylor, 2015; Morehead et al., 2017; Taylor & Ivry, 2011). Interestingly, this confound has also led to several misleading interpretations regarding motor adaptation (Bond & Taylor, 2015; McDougle et al., 2015), for instance with regards to savings (Morehead et al., 2015), prompting a much-needed reconsideration of the current literature regarding reinforcement as well.

In one of the earliest works suggesting a contribution of reinforcement in motor adaptation, it was demonstrated that repetition of a large visuomotor rotation leads to a strengthening of the associated visuomotor memory (V. S. Huang et al., 2011). While the results observed could be explained by an implicit reinforcement interpretation, it could also be explained by an entirely explicit framework (Bond & Taylor, 2015; Taylor & Ivry, 2011): since repetition was induced by assigning the displacements so the absolute reach angle accounting for them were all identical (see figure 1.6), participants may have explicitly noted and remembered this characteristic and simply repeated it during re-learning, leading to the observed faster re-learning rates. Other work from the same year shows that introduction of gradual binary feedback does lead to some degree of awareness, as measured by proprioceptive shift (Izawa & Shadmehr, 2011), in line with our results from chapters 2-3. More recently, a report that monetary reward enhances retention of a visuomotor memory (Galea et al., 2015) might be explained in the framework of explicit control. In that study, participants were rewarded for reaching accurately to targets in a visuomotor rotation task. In the light of recent work on explicit control (Bond & Taylor, 2015; Taylor & Ivry, 2011), this implies that explicitly re-aiming off target to hit them was rewarded monetarily, especially early in the task when cerebellar contribution was weak

(Huberdeau et al., 2015). Therefore, during a no-feedback retention block following the rewarded adaptation block, participants may consider it more financially productive to keep re-aiming off target volitionally in case of a hidden reward region. Finally, electroencephalography recordings during a visuomotor adaptation task (Palidis et al., 2019) exhibited reward-related event-related potentials that may suggest an explicit processing of information (Loonis et al., 2017), in line with the thesis of this work.

Other works lead to a more equivocal picture. Small and gradually introduced perturbations are usually associated with less awareness (Christou et al., 2016), meaning that an associated reward effect can hardly be considered within an explicit framework. Notably, it has been shown that four degrees reward-signalled perturbations tend to lead to small but consistent tuning of the mean reach angle, an effect not observed in PD patients (Pekny et al., 2015). Because that study does not use sensory prediction error, this suggests that reward can shape motor behaviour, at least on a small scale (four degrees). In another study employing sensory prediction error, manipulating target sizes shows that motor adaptation stops when the cursor is brought back to the target regardless of the target size, suggesting that a rewarding outcome (target "hits") can stop error-based learning independently of the absolute value of adaptation plateau (H. E. Kim et al., 2019). In line, artificially inflating the reward rate (target "hits") also reduces the plateau value of the adaptation curve (Leow et al., 2018), though this last study does not control for explicit contribution.

Altogether, these studies suggest that reward plays a role in shaping motor behaviour implicitly and possibly modulating implicit cerebellar adaptation, but the emerging picture is that its impact reduces to fine-tuning rather than wider effects such as enhanced retention or increased learning rates, and several reports should be re-examined accordingly. At the very least, because reinforcement and explicit control both rely on task success to learn, monitoring for possible confounds is essential. Furthermore, the fact that they rely on the same feedback may be taken as theoretical evidence that explicit control is simply the counterpart of model-free reinforcement, that is, model-based reinforcement.

6.3 Implications of this work for rehabilitation and improving training procedures

The goal of reinforcement-based motor learning research is to design enhanced rehabilitation and training procedures for clinical populations like stroke patients and healthy populations like sports and artists performers. In these regards, the results shown here suggest that explicitly grounded methods such as coaching may be more productive and impactful. Nevertheless, future advances toward this goal lie either in (1) a re-assessment of reinforcement on fully implicit motor learning or (2) on how to promote a transfer (consolidation) from explicit procedural memory to an implicit component.

One way to achieve (1) is using small and gradual perturbations to avoid conflation with an explicit component (Christou et al., 2016). Closed-loop designs, where the rewarded action is a function of individual performance similar to that used in chapters 4-5 might also lead to implicit learning, by avoiding artificial, enforced constraints on motor output demands (Holland et al., 2018; Therrien et al., 2016). Finally, using demanding dual tasks to restrict explicit contributions might promote implicit learning (Holland et al., 2018). Achieving (2) may require longer training sessions, or likely, multi-day practice sessions (Mulavara et al., 2010; Reis et al., 2009; Ruitenberg et al., 2018), two possibilities that may be tested in the future. However, there is little evidence to date that any form of error-based learning such as cerebellar learning results in a long-lasting memory as opposed to more complex skill learning tasks, and therefore, it appears more promising to focus on such complex tasks to unravel a practically applicable effect of reinforcement. One possible effect of reward that has not been assessed yet is in consolidation of motor memory after training sessions, known as offline gains (Reis et al., 2009). In line, motor memories are sensitive to offline gains, and other forms of emotional memories have already been shown to undergo offline processing as well (Walker & van der Helm, 2009).

Regarding clinical rehabilitation, some reports of reward-driven enhancements of motor function in stroke patients have been reported (Goodman et al., 2014; Quattrocchi et al., 2017), but this may be more related to a motivational effect, increasing attention span through "gamification" of the task (van der Kooij et al., 2019). We discuss the relationship between attention and reward further in section 6.5.

6.4 The interaction of reward and explicit control in complex motor skills

By definition, a complex task involves linking together a set of simpler tasks into a co-ordinated movement with a higher-order goal, such as driving a car or putting a golf ball in its hole with a club (Maxwell et al., 2001). Considering more intricate motor skills has a strong theoretical advantage over simpler tasks such as reaching movements, in that it includes explicit control by essence (Ghilardi et al., 2009; Robertson, 2007; Stanley & Krakauer, 2013) and it is therefore unnecessary to control for it. However, it remains possible, and perhaps surprisingly beneficial to reduce its contribution (*e.g.* "analogy learning"; Maxwell et al., 2001), as several studies have shown that this strengthens its transferability (Liao & Masters, 2001) and reduces susceptibility to potential distractors (Maxwell et al., 2001) or to stress during performance (Buszard & Masters, 2018; Maxwell et al., 2000). Other works demonstrate that reward schedule impacts skill acquisition as well (E. Dayan et al., 2014; Maxwell et al., 2001; Hamel et al., 2019). Interestingly, one of these studies achieves better learning by minimizing the amount of errors (punishment) to avoid triggering reflective and deduction-based learning, that is, explicit control (Maxwell et al., 2001). In contrast, a study on astronauts re-adapting to earth gravity after long-duration spaceflights shows that between-days consolidation levels correlate with early, likely explicit learning (Mulavara et al., 2010). Overall, explicit control appears to have a significant impact on complex skill learning, though how this occurs is still a puzzling issue, as different works suggest that it is either beneficial or disadvantageous.

Some evidence may be found when looking at long-term improvements over hundreds of hours of practice (Gray & Lindstedt, 2017). At this timescale, individual performance profiles show that improvements actually occur in steps, with a performance plateau being usually bounded by short periods of great improvements called "leaps" (Gallistel et al., 2004). These leaps indicate hierarchical reorganization of the skill policy, or put in our terms, structural learning and explorative behaviour—which is explicit control. Therefore, explicit control is clearly beneficial in a timely manner, and since manipulating error rates can exacerbate or reduce its contribution (Maxwell et al., 2001), reinforcement has great potential in this regard. One could imagine an enhanced closed-loop reward function which tracks performance history, and diminishes reward when performance stagnates at a suboptimal plateau to trigger explorative behaviour and a leap forward. Such "smart" reward function should be calibrated to take into account difficulty and avoid "choking" behaviour (individuals giving up on a task because positive feedback is scarce or non-existent) as exhibited by some individuals in chapter 2-3 and other studies (van der Kooij et al., 2018; Manley et al., 2014). This possibility may be tested using a complex task that displays this sort of learning curve such as typing, and assessing when the introduction of reward promotes transitions from plateaus to "leaps" (Gray & Lindstedt, 2017). However, it remains unclear how efficient this method would be compared to instruction-based coaching methods for instance, such as the popular "deliberate practice" method (Ericsson et al., 1993). Deliberate practice relies on explicitly given instructions, individualised tutoring, diagnosing of errors and customised training programs to meet pre-specified performance sub-goals. One appealing prospect of an alternative, reward-based method is that it would not rely on individual tutoring sessions, permitting a large-scale, formalised application of this concept.

6.5 Reinforcement as attentional drive

In chapter 5, we hypothesized that selection improvements with reward may be driven by cholinergic input to the visual areas. This hypothesis bears conceptual implications, as it also implies that reward-based effects may in fact be attentional effects (Sarter et al., 2005). For instance, improvements in skill learning may simply rely on enhanced encoding of working memory items, leading to a stronger consolidation and better offline gains, and eventually an enhanced motor memory. Stronger visual feedback control during rewarded reaching movements may also be enabled through increased cholinergic modulation of occipital regions (Goard & Dan, 2009; Pinto et al., 2013), enhancing visual information representation and therefore permitting higher gains. Overall, there is a rich literature discussing the impact of reward on attentional processing (Anderson, 2016; Chelazzi et al., 2013; Maunsell, 2004; Sawaki et al., 2015), and while most of it is out of scope for this thesis, it could potentially explain some of the findings reported here. Finally, it is unlikely that a dissociation between attentional drive and reinforcement drive is a fruitful debate (Maunsell, 2004). Regarding multi-day learning for instance (Reis et al., 2009; Ruitenberg et al., 2018; Telgen et al., 2014), it is indeed irrelevant whether the reward-driven improvements in learning are due to reward per se or to an attentional drive triggered by reward that would result in a better encoding of information (Anderson, 2016; Chelazzi et al., 2013; Maunsell, 2004). Even if it was the case, attention is fundamentally difficult to manipulate and exploiting reward in that end remains worthwhile.

6.6 Genetics and motor learning research

While great effort has been put into dopamine-related genetics research (Berke, 2018), considering the possible implication of attention in reward-based effects may incite to consider other neuromodulators involved in arousal and sleep regulation. Notably, serotonin and acetylcholine have been known to have a central role in these mechanisms (Quist et al., 2003; Sarter et al., 2005). There is currently little research investigating the role of individual genetic variation of these neuromodulators in motor learning.

On the other hand, results observed in chapter 3 provide a cautionary tale to any future genetic investigation in motor control. One may argue that motor control appears to rely on a complex network englobing a vast set of brain regions (Shadmehr & Krakauer, 2008); consequently, one single gene, or a reduced set of genes cannot be linked to a specific motor function. The situation is radically different regarding reinforcement and decision-making, with the neuromodulator dopamine sitting at the core of most theories on the matter (Berke, 2018), leading to an abundant literature on dopamine-related genes in that field. However, this doesn't nullify the complexity of the motor control system beyond reinforcement, which is very likely to dilute any genetic effect rooted in individual variability of dopamine-related genes in healthy population. Indeed, the natural redundancy the motor system displays through its complexity makes it prone to compensatory behaviour. On the other hand, effects may still be observed in pathological cases such as PD patients because of the scale of the dopaminergic system impairment (Manohar et al., 2015; Mazzoni et al., 2007; Pekny et al., 2015). Finally, although we observe no relationship between dopaminergic gene profiles and reward-based learning performance, this evidence is not causal, and manipulating the dopaminergic state of participants may provide additional information regarding the involvement of dopamine in reward-based motor learning. Pharmacological manipulations of the neuro-modulatory state using *e.g.* L-DOPA may allow assessing a possible causal rather than correlational relationship.

6.7 Stiffness and reward in more complex tasks

In chapter 4 we ask why enhanced motor performance is supported by different control strategies when considering different plants such as eye muscles or the upper limb. This can only be addressed through a quantitative approach, assessing what is the underlying cost of increased feedback gain or stiffness levels of different motor plants by careful modelling of the motor system with stronger biological constraints (Bhushan & Shadmehr, 1999). Such modelling will also allow simulations to assess what is the optimal time course of these manipulations along a reach, possibly confirming the hypothesis we propose that optimal stiffness control requires higher stiffness at the vicinity of the target than during the early part of the reach.

This question can also be extended to chunking (Sosnik et al., 2004). While high stiffness is beneficial against perturbations or as a "braking" strategy to cancel for effector inertia near a target position, it can also hinder smooth transitions between movement subsets such as observed in coarticulation, which is a central element of chunking performance (Sosnik et al., 2004). Since reward has a significant impact on sequence learning (Wachter et al., 2009) and current preliminary work shows that acquisition and retention of chunking improves with reward, it may be that stiffness will be reduced with reward during transitions. Again, this matter may be addressed in future work where stiffness could be measured using a variation of the displacement technique employed in chapter 4-5 during a multi-target reaching task that results in co-articulation of movement.

6.8 Methodological approach

There is currently a much-heated debate over the methodology underlying the scientific process. Notably, the use of p-values has been strongly criticised (Cumming, 2014; Lakens et al., 2018), the absence of replicability of even some of the most classical works (Baker, 2016; Open Science Collaboration, 2015), and presumably widespread HARKing practices (Bosco et al., 2016) are elements that steered a great deal of attention. Fortunately, methodological amendments have been formulated, and much effort has been put to implement them in the most sensible way possible in the work presented here.

As more complex experiments tend to be more exposed to HARKing, methodologically heavier parts of the thesis (chapter 3 and 5 have been pre-registered using Open Science Framework. While ideally, one may perform pilot experiments, draw conclusions *post-hoc* and put those conclusions to the test through pre-registered replication attempts, that research model is often made unviable by productivity and funding demands that do not allow for the consequent time and financial expenditure. For this reason, while chapter 2 and 4 are designed and were published as their own standalone projects, they also served as the basis for preregistration of chapter 3 and 5. I believe this approach to be acceptable, granted one remains thoroughly clear about what was predicted and what was concluded *post-hoc* from the data. Chapter 4 is a good instance of this point, because it contains a large set of experiments (4 experiments and 1 modelling section), where the logical flow was made as transparent as possible, while using each follow-up experiment as a qualitative replication of previous results. In that regard, chapter 2-3 also contain several within-study replications of proposed results, ascertaining the reliability of the conclusions. Finally, all published work was made available free of charge using a pre-print server, and datasets and analysis scripts were also all made available online for anyone to download as they please without having to request it formally. This was applied to support the Open-Access initiative.

Regarding statistical methods, much care was given to use a large participant base, if possible using a minimum of 20 participant per experimental group. We attempted to move away from standard error of the mean as a basis for error estimation to a bootstrapped 95% CI, as this was shown to be a much more informative method to complement the null-hypothesis significance testing (NHST) methodology (Cumming, 2014). Statistics experts have even held a more extreme view, suggesting leaving out the NHST approach altogether for CIs estimation (Cumming, 2014), but this being still a debated matter, we adopted a more conservative approach by reporting p-values nonetheless. Others have suggested not constraining analyses to NHST and rather rely on a careful observation of underlying data distribution (Rousselet et al., 2016, 2017), which we followed by showing all individual dots in group analyses and adding group distributions where possible.

6.9 Conclusions

Over the past several years, it became increasingly clear that reward has a significant impact on motor control and on the course of motor learning. In this thesis, the underlying nature for this phenomenon is investigated. It was observed that reinforcement, which was taken as an implicitly driven phenomenon is rooted in explicit control, prompting a re-appraisal of the literature and of its potential for enhanced rehabilitation and training procedures. However, this does not discard reinforcement as a mere confound, but rather asks to re-centre the debate from the fallacious view of fundamentally implicit skill learning to a more complex understanding of reinforcement-driven skill learning that acknowledges the central role of explicit control and structural learning. In this regard this work is in full accordance with the current shift observed in the field toward a more integrated view of strategic control with more implicit aspects, such as for cerebellar adaptation (Bond & Taylor, 2015), sequence learning (Robertson, 2007) and complex skill learning (Stanley & Krakauer, 2013). Finally, chapters 4-5 consider the effect of reward on motor control, and how reward supports enhanced control mechanistically. It is demonstrated that while there is evidence of increased feedback control during reaching under reward, in line with the recent literature, another mechanism allows for this enhanced control in the form of stiffness regulation.

References

1000 Genomes Project Consortium, T., Gibbs, R. A., Boerwinkle, E., Doddapaneni, H., Han, Y., Korchina, V., . . . et al. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. doi: 10.1038/nature15393

Altman, D. G. & Bland, J. M. (1995). Absence of evidence is not evidence of absence. *British Medical Journal*, *311*(7003), 485.

Alvarez-Icaza, R. & Boahen, K. (2012). Inferior olive mirrors joint dynamics to implement an inverse controller. , *106*(8), 429–439. doi: 10.1007/s00422-012-0498-2

Ames, K. C., Ryu, S. I. & Shenoy, K. V. (2019). Simultaneous motor preparation and execution in a last-moment reach correction task. *Nature Communications*, *10*(1), 2718. doi: 10.1038/s41467-019-10772-2

Anderson, B. A. (2016). The attention habit: how reward learning shapes attentional selection: The attention habit. *Annals of the New York Academy of Sciences*, *1369*(1), 24–39. doi: 10.1111/nyas.12957

Anguera, J. A., Bernard, J. A., Jaeggi, S. M., Buschkuehl, M., Benson, B. L., Jennett, S., . . . Seidler, R. D. (2012). The effects of working memory resource depletion and training on sensorimotor adaptation. *Behavioural Brain Research*, *228*(1), 107–115. doi: 10.1016/j.bbr.2011.11.040

Anguera, J. A., Reuter-Lorenz, P. A., Willingham, D. T. & Seidler, R. D. (2010). Contributions of spatial working memory to visuomotor learning. *Journal of cognitive neuroscience*, *22*(9), 1917–1930.

Arai, N., Lu, M.-K., Ugawa, Y. & Ziemann, U. (2012). Effective connectivity between human supplementary motor area and primary motor cortex: a paired-coil tms study. *Experimental Brain Research*, *220*(1), 79–87. doi: 10.1007/s00221-012-3117-5

Arai, N., Muller-Dahlhaus, F., Murakami, T., Bliem, B., Lu, M.-K., Ugawa, Y. & Ziemann, U. (2011). State-dependent and timing-dependent bidirectional associative plasticity in the human sma-m1 network. *Journal of Neuroscience*, *31*(43), 15376–15383. doi: 10.1523/JNEUROSCI.2271-11.2011

Arnsten, A. F. (1998). Catecholamine modulation of prefrontal cortical cognitive function. *Trends in Cognitive Sciences*, *2*(11), 436–447. doi: 10.1016/S1364-6613(98)01240-6

Baker, M. (2016). A nature survey lifts the lid on how researchers view the ‘crisis’ rocking science and what they think will help. *Nature*, *533*(7604), 452–454. doi: 10.1038/533452a

Balleine, B. W. & O’Doherty, J. P. (2010). Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, *35*(1), 48–69. doi: 10.1038/npp.2009.131

Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Becker, M. I. & Person, A. L. (2019). Cerebellar control of reach kinematics for endpoint precision. *Neuron*, *103*(2), 335-348. doi: 10.1016/j.neuron.2019.05.007

Benson, B. L., Anguera, J. A. & Seidler, R. D. (2011). A spatial explicit strategy reduces error but interferes with sensorimotor adaptation. *Journal of Neurophysiology*, *105*(6), 2843–2851. doi: 10.1152/jn.00002.2011

Berke, J. D. (2018). What does dopamine mean? *Nature Neuroscience*, *21*(6), 787–793. doi: 10.1038/s41593-018-0152-y

Berret, B., Castanier, C., Bastide, S. & Deroche, T. (2018). Vigour of self-paced reaching movement: cost of time and individual traits. *Scientific Reports*, *8*(1), 10655. doi: 10.1038/s41598-018-28979-6

Beschin, N., Denis, M., Logie, R. H. & Della Sala, S. (2005). Dissociating mental transformations and visuo-spatial storage in working memory: Evidence from representational neglect. *Memory*, *13*(3–4), 430–434. doi: 10.1080/09658210344000431

Bhushan, N. (1998). *A computational approach to human adaptive motor control* (Unpublished doctoral dissertation). Johns Hopkins University.

Bhushan, N. & Shadmehr, R. (1999). Computational nature of human adaptive control during learning of reaching movements in force ®elds. *Biological Cybernetics*, *81*(1), 39–60. doi: 10.1007/s004220050543

Blair, K. S., Otero, M., Teng, C., Jacobs, M., Odenheimer, S., Pine, D. S. & Blair, R. (2013). Dissociable roles of ventromedial prefrontal cortex (vmpfc) and rostral anterior cingulate cortex (racc) in value representation and optimistic bias. *NeuroImage*, *78*, 103–110. doi: 10.1016/j.neuroimage.2013.03.063

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

Bosco, F. A., Aguinis, H., Field, J. G., Pierce, C. A. & Dalton, D. R. (2016). Harking's threat to organizational research: Evidence from primary and meta-analytic sources. *Personnel Psychology*, *69*(3), 709–750. doi: 10.1111/peps.12111

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Brennan, A. E. & Smith, M. A. (2015). The decay of motor memories is independent of context change detection. *PLOS Computational Biology*, *11*(6), e1004278. doi: 10.1371/journal.pcbi.1004278

Bundt, C., Abrahamse, E. L., Braem, S., Brass, M. & Notebaert, W. (2016). Reward anticipation modulates primary motor cortex excitability during task preparation. *NeuroImage*, *142*, 483–488. doi: 10.1016/j.neuroimage.2016.07.013

Burdet, E., Osu, R., Franklin, D. W., Yoshioka, T., Milner, T. E. & Kawato, M. (2000). A method for measuring endpoint stiffness during multi-joint arm movements. *Journal of Biomechanics*, 5.

Buszard, T. & Masters, R. S. W. (2018). Adapting, correcting and sequencing movements: does working-memory capacity play a role? *International Review of Sport and Exercise Psychology*, *11*(1), 258–278. doi: 10.1080/1750984X.2017.1323940

Bütefisch, C. M., Davis, B. C., Wise, S. P., Sawaki, L., Kopylev, L., Classen, J. & Cohen, L. G. (2000). Mechanisms of use-dependent plasticity in the human motor cortex. *Proceedings of the national academy of sciences*, *97*(7), 3661–3665.

Carroll, T. J., McNamee, D., Ingram, J. N. & Wolpert, D. M. (2019). Rapid visuomotor responses reflect value-based decisions. *Journal of Neuroscience*, *39*(20), 3906–3920. doi: 10.1523/JNEUROSCI.1934-18.2019

Castiello, U. (2005). The neuroscience of grasping. *Nature Reviews Neuroscience*, *6*(9), 726–736. doi: 10.1038/nrn1744

Chelazzi, L., Perlato, A., Santandrea, E. & Della Libera, C. (2013). Rewards teach visual selective attention. *Vision Research*, *85*, 58–72. doi: 10.1016/j.visres.2012.12.005

Chen, X., Holland, P. & Galea, J. M. (2018). The effects of reward and punishment on motor skill learning. *Current Opinion in Behavioral Sciences*, *20*, 83–88. doi: 10.1016/j.cobeha.2017.11.011

Chen, X., Mohr, K. & Galea, J. M. (2017a). Predicting explorative motor learning using decision-making and motor noise. *PLOS Computational Biology*, *13*(4), e1005503. doi: 10.1371/journal.pcbi.1005503

Chen, X., Mohr, K. & Galea, J. M. (2017b). Predicting explorative motor learning using decision-making and motor noise. *PLOS Computational Biology*, *13*(4), e1005503. doi: 10.1371/journal.pcbi.1005503

Cheng, S. & Sabes, P. N. (2006). Modeling sensorimotor learning with linear dynamical systems. *Neural computation*, *18*(4), 760–793.

Chen-Harris, H., Joiner, W. M., Ethier, V., Zee, D. S. & Shadmehr, R. (2008). Adaptive control of saccades via internal feedback. *Journal of Neuroscience*, *28*(11), 2804-2813. doi: 10.1523/JNEUROSCI.5300-07.2008

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

Churchland, M. M., Afshar, A. & Shenoy, K. V. (2006). A central source of movement variability. *Neuron*, *52*(6), 1085–1096. doi: 10.1016/j.neuron.2006.10.034

Classen, J., Liepert, J., Wise, S. P., Hallett, M. & Cohen, L. G. (1998). Rapid plasticity of human cortical movement representation induced by practice. *Journal of neurophysiology*, *79*(2), 1117–1123.

Codol, O., Holland, P. J. & Galea, J. M. (2018). The relationship between reinforcement and explicit control during visuomotor adaptation. *Scientific Reports*, *8*(1). doi: 10.1038/s41598-018-27378-1

Cohen, A. L., Sanborn, A. N. & Shiffrin, R. M. (2008). Model evaluation using grouped or individual data. *Psychonomic Bulletin & Review*, *15*(4), 692-712. doi: 10.3758/PBR.15.4.692

Cohen, M. S., Kosslyn, S. M., Breiter, H. C., DiGirolamo, G. J., Thompson, W. L., Anderson, A. K., … Belliveau, J. W. (1996). Changes in cortical activity during mental rotation a mapping study using functional mri. *Brain*, *119*(1), 89–100. doi: 10.1093/brain/119.1.89

Crevecoeur, F., Kurtzer, I., Bourke, T. & Scott, S. H. (2013). Feedback responses rapidly scale with the urgency to correct for external perturbations. *Journal of Neurophysiology*, *110*(6), 1323–1332. doi: 10.1152/jn.00216.2013

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. doi: 10.1177/0956797613504966

Dash, P. K., Moore, A. N., Kobori, N. & Runyan, J. D. (2007). Molecular activity underlying working memory. *Learning & Memory*, *14*(8), 554–563. doi: 10.1101/lm.558707

da Silva, J. A., Tecuapetla, F., Paixão, V. & Costa, R. M. (2018). Dopamine neuron activity before action initiation gates and invigorates future movements. *Nature*, *554*(7691), 244–248. doi: 10.1038/nature25457

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215. doi: 10.1016/j.neuron.2011.02.027

Daw, N. D., Niv, Y. & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711. doi: 10.1038/nn1560

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879. doi: 10.1038/nature04766

Dayan, E., Averbeck, B. B., Richmond, B. J. & Cohen, L. G. (2014). Stochastic reinforcement benefits skill acquisition. *Learning & Memory*, *21*(3), 140–142. doi: 10.1101/lm.032417.113

Dayan, P. & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(4), 429–453. doi: 10.3758/CABN.8.4.429

Deserno, L., Huys, Q. J. M., Boehme, R., Buchert, R., Heinze, H.-J., Grace, A. A., ... Schlagenhauf, F. (2015). Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. *Proceedings of the National Academy of Sciences*, *112*(5), 1595–1600. doi: 10.1073/pnas.1417219112

De Zeeuw, C. I. & Ten Brinke, M. M. (2015). Motor learning and the cerebellum. *Cold Spring Harbor Perspectives in Biology*, *7*(9), a021683. doi: 10.1101/cshperspect.a021683

Dideriksen, J. L., Negro, F., Enoka, R. M. & Farina, D. (2012). Motor unit recruitment strategies and muscle properties determine the influence of synaptic noise on force steadiness. *Journal of Neurophysiology*, *107*(12), 3357–3369. doi: 10.1152/jn.00938.2011

Diedrichsen, J. & Kornysheva, K. (2015). Motor skill learning between selection and execution. *Trends in Cognitive Neuroscience*, *19*(4), 227-233. doi: 10.1016/j.tics.2015.02.003

Dolan, R. J. & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, *80*(2), 312–325. doi: 10.1016/j.neuron.2013.09.007

Doll, B. B., Bath, K. G., Daw, N. D. & Frank, M. J. (2016). Variability in dopamine genes dissociates model-based and model-free reinforcement learning. *Journal of Neuroscience*, *36*(4), 1211–1222. doi: 10.1523/JNEUROSCI.1901-15.2016

Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D. & Daw, N. D. (2015). Model-based choices involve prospective neural activity. *Nature Neuroscience*, *18*(5), 767–772. doi: 10.1038/nn.3981

Douskos, C. (2017). The spontaneousness of skill and the impulsivity of habit. *Synthese*. doi: 10.1007/s11229-017-1658-7

Egan, M. F., Goldberg, T. E., Kolachana, B. S., Callicott, J. H., Mazzanti, C. M., Straub, R. E., ... Weinberger, D. R. (2001). Effect of comt val108/158 met genotype on frontal lobe function and risk for schizophrenia. *Proceedings of the National Academy of Sciences*, *98*(12), 6917–6922.

Ericsson, K. A., Krampe, R. T. & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406.

Fernandez-Ruiz, J., Wong, W., Armstrong, I. T. & Flanagan, J. R. (2011). Relation between reaction time and reach errors during visuomotor adaptation. *Behavioural Brain Research*, *219*(1), 8–14. doi: 10.1016/j.bbr.2010.11.060

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*(6), 381–391. doi: 10.1037/h0055392

Floeter, M. K. (2010). Structure and function of muscle fibers and motor units. In G. Karpati, D. Hilton-Jones, K. Bushby & R. C. Griggs (Eds.), *Disorders of* *voluntary muscle* (8th ed., p. 1–19). Cambridge University Press. doi: 10.1017/CBO9780511674747.005

Frank, M. J., Doll, B. B., Oas-Terpstra, J. & Moreno, F. (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature Neuroscience*, *12*(8), 1062–1068. doi: 10.1038/nn.2342

Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T. & Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences*, *104*(41), 16311–16316.

Franklin, D. W., Liaw, G., Milner, T. E., Osu, R., Burdet, E. & Kawato, M. (2007). Endpoint stiffness of the arm is directionally tuned to instability in the environment. *Journal of Neuroscience*, *27*(29), 7705–7716. doi: 10.1523/JNEUROSCI.0968-07.2007

Franklin, D. W., Osu, R., Burdet, E., Kawato, M. & Milner, T. E. (2003). Adaptation to stable and unstable dynamics achieved by combined impedance control and inverse dynamics model. *Journal of Neurophysiology*, *90*(5), 3270–3282. doi: 10.1152/jn.01112.2002

Galaro, J. K., Celnik, P. & Chib, V. S. (2019). Motor cortex excitability reflects the subjective value of reward and mediates its effects on incentive-motivated performance. *The Journal of Neuroscience*, *39*(7), 1236–1248. doi: 10.1523/JNEUROSCI.1254-18.2018

Galea, J. M., Albert, N. B., Ditye, T. & Miall, R. C. (2010). Disruption of the dorsolateral prefrontal cortex facilitates the consolidation of procedural skills. *Journal of Cognitive Neuroscience*, *22*(6), 1158–1164. doi: 10.1162/jocn.2009.21259

Galea, J. M., Mallia, E., Rothwell, J. & Diedrichsen, J. (2015). The dissociable effects of punishment and reward on motor learning. *Nature Neuroscience*, *18*(4), 597–602. doi: 10.1038/nn.3956

Galea, J. M., Vazquez, A., Pasricha, N., Orban de Xivry, J.-J. & Celnik, P. (2011). Dissociating the roles of the cerebellum and motor cortex during adaptive learning: The motor cortex retains what the cerebellum learns. *Cerebral Cortex*, *21*(8), 1761–1770. doi: 10.1093/cercor/bhq246

Gallistel, C. R., Fairhurst, S. & Balsam, P. (2004). The learning curve: Implications of a quantitative analysis. *Proceedings of the National Academy of Sciences*, *101*(36), 13124–13131. doi: 10.1073/pnas.0404965101

Gershman, S. J. & Schoenbaum, G. (2017). Rethinking dopamine prediction errors. *bioRxiv*, 239731.

Ghilardi, M. F., Moisello, C., Silvestri, G., Ghez, C. & Krakauer, J. W. (2009). Learning of a sequential motor skill comprises explicit and implicit components that consolidate differently. *Journal of Neurophysiology*, *101*(5), 2218–2229. doi: 10.1152/jn.01138.2007

Gläscher, J., Daw, N., Dayan, P. & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*(4), 585–595. doi: 10.1016/j.neuron.2010.04.016

Goard, M. & Dan, Y. (2009). Basal forebrain activation enhances cortical coding of natural scenes. *Nature Neuroscience*, *12*(11), 1444–1449. doi: 10.1038/nn.2402

Goldberg, T. E., Egan, M. F., Gscheidle, T., Coppola, R., Weickert, T., Kolachana, B. S., ... Weinberger, D. R. (2003). Executive subprocesses in working memory: Relationship to catechol-o-methyltransferase val158met genotype and schizophrenia. *Archives of General Psychiatry*, *60*(9), 889–896. doi: 10.1001/archpsyc.60.9.889

Gomi, H. & Kawato, M. (1993). Neural network control for a closed-loop system using feedback-error-learning. *Neural Networks*, *6*(7), 933–946.

Goodman, R. N., Rietschel, J. C., Roy, A., Jung, B. C., Diaz, J., Macko, R. F. & Forrester, L. W. (2014). Increased reward in ankle robotics training enhances motor control and cortical efficiency in stroke. *Journal of Rehabilitation Research and Development*, *51*(2), 213–228. doi: 10.1682/JRRD.2013.02.0050

Gray, W. D. & Lindstedt, J. K. (2017). Plateaus, dips, and leaps: Where to look for inventions and discoveries during skilled performance. *Cognitive Science*, *41*(7), 1838–1870. doi: 10.1111/cogs.12412

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annual Review of Neuroscience*, *31*(1), 359–387. doi: 10.1146/annurev.neuro.29.051605.112851

Gribble, P. L., Mullin, L. I., Cothros, N. & Mattar, A. (2003). Role of cocontraction in arm movement accuracy. *Journal of Neurophysiology*, *89*(5), 2396–2405. doi: 10.1152/jn.01020.2002

Gurney, K., Prescott, T. J. & Redgrave, P. (2001). A computational model of action selection in the basal ganglia. i. a new functional anatomy. *Biological Cybernetics*, *84*(6), 401–410. doi: 10.1007/PL00007984

Haith, A. M., Huberdeau, D. M. & Krakauer, J. W. (2015). The influence of movement preparation time on the expression of visuomotor learning and savings. *Journal of Neuroscience*, *35*(13), 5109–5117. doi: 10.1523/JNEUROSCI.3869-14.2015

Haith, A. M. & Krakauer, J. W. (2013). Model-based and model-free mechanisms of human motor learning. In M. J. Richardson, M. A. Riley & K. Shockley (Eds.), *Progress in motor control* (Vol. 782, p. 1–21). Springer New York.

Haith, A. M., Pakpoor, J. & Krakauer, J. W. (2016). Independence of movement preparation and movement initiation. *Journal of Neuroscience*, *36*(10), 3007–3015. doi: 10.1523/JNEUROSCI.3245-15.2016

Hamann, S. B., Ely, T. D., Grafton, S. T. & Kilts, C. D. (1999). Amygdala activity related to enhanced memory for pleasant and aversive stimuli. *Nature Neuroscience*, *2*(3), 289–293.

Hamel, R., Côté, K., Matte, A., Lepage, J.-F. & Bernier, P.-M. (2019). Rewards interact with repetition-dependent learning toenhance long-term retention of motor memories. *Annals of the New York Academy of Sciences*, *In Press*. doi: 10.1111/nyas.14171

Hamel, R., Savoie, F.-A., Lacroix, A., Whittingstall, K., Trempe, M. & Bernier, P.-M. (2018). Added value of money on motor performance feedback: Increased left central beta-band power for rewards and fronto-central theta-band power for punishments. *NeuroImage*, *179*, 63–78. doi: 10.1016/j.neuroimage.2018.06.032

Hamilton, A. F., Jones, K. E. & Wolpert, D. M. (2004). The scaling of motor noise with muscle strength and motor unit number in humans. *Experimental Brain Research*, *157*(4), 417–430. doi: 10.1007/s00221-004-1856-7

Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E., ... Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, *6*, e4794. doi: 10.7717/peerj.4794

Haruno, M., Wolpert, D. M. & Kawato, M. (2001). Mosaic model for sensorimotor learning and control. *Neural Computation*, *13*(10), 2201–2220.

Hendrix, C. M., Campbell, B. A., Tittle, B. J., Johnson, L. A., Baker, K. B., Johnson, M. D., ... Vitek, J. L. (2018). Predictive encoding of motor behavior in the supplementary motor area is disrupted in parkinsonism. *Journal of Neurophysiology*, *120*(3), 1247–1255. doi: 10.1152/jn.00306.2018

Hikosaka, O., Nakamura, K., Sakai, K. & Nakahara, H. (2002). Central mechanisms of motor skill learning. *Current Opinion in Neurobiology*, *12*(2), 217–222. doi: 10.1016/S0959-4388(02)00307-0

Hirashima, M. & Nozaki, D. (2012). Learning with slight forgetting optimizes sensorimotor transformation in redundant motor systems. *PLoS Computational Biology*, *8*(6), e1002590. doi: 10.1371/journal.pcbi.1002590

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

Holmes, G. (1939). The cerebellum of man. *Brain*, *62*(1), 1–30.

Honda, T., Nagao, S., Hashimoto, Y., Ishikawa, K., Yokota, T., Mizusawa, H. & Ito, M. (2018). Tandem internal models execute motor learning in the cerebellum. , *115*(28), 7428–7433. doi: 10.1073/pnas.1716489115

Huang, V. S., Haith, A., Mazzoni, P. & Krakauer, J. W. (2011). Rethinking motor learning and savings in adaptation paradigms: Model-free memory for successful actions combines with internal models. *Neuron*, *70*(4), 787–801. doi: 10.1016/j.neuron.2011.04.012

Huang, Y.-Z., Edwards, M. J., Rounis, E., Bhatia, K. P. & Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron*, *45*(2), 201–206. doi: 10.1016/j.neuron.2004.12.033

Huberdeau, D. M., Krakauer, J. W. & Haith, A. M. (2015). Dual-process decomposition in human sensorimotor adaptation. *Current Opinion in Neurobiology*, *33*, 71–77. doi: 10.1016/j.conb.2015.03.003

Hunt, L. T. & Hayden, B. Y. (2017). A distributed, hierarchical and recurrent framework for reward-based choice. *Nature Reviews Neuroscience*, *18*(3), 172–182. doi: 10.1038/nrn.2017.7

Hutter, S. A. & Taylor, J. A. (2018). Relative sensitivity of explicit reaiming and implicit motor adaptation. *Journal of Neurophysiology*, *120*(5), 2640-2648. doi: 10.1152/jn.00283.2018

Huys, Q. J. M., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P. & Roiser, J. P. (2012). Bonsai trees in your head: How the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Computational Biology*, *8*(3), e1002410. doi: 10.1371/journal.pcbi.1002410

Huys, Q. J. M., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., . . . Roiser, J. P. (2015). Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences*, *112*(10), 3098–3103. doi: 10.1073/pnas.1414219112

Hwang, E. J., Smith, M. A. & Shadmehr, R. (2006). Dissociable effects of the implicit and explicit memory systems on learning control of reaching. *Experimental Brain Research*, *173*(3), 425–437. doi: 10.1007/s00221-006-0391-0

Hübner, R. & Schlösser, J. (2010). Monetary reward increases attentional effort in the flanker task. *Psychonomic Bulletin & Review*, *17*(6), 821–826. doi: 10.3758/PBR.17.6.821

Ikegami, T., Hirashima, M., Osu, R. & Nozaki, D. (2012). Intermittent visual feedback can boost motor learning of rhythmic movements: Evidence for error feedback beyond cycles. *Journal of Neuroscience*, *32*(2), 653-657. doi: 10.1523/JNEUROSCI.4230-11.2012

Ikegami, T., Hirashima, M., Taga, G. & Nozaki, D. (2010). Asymmetric transfer of visuomotor learning between discrete and rhythmic movements. *Journal of Neuroscience*, *30*(12), 4515-4521. doi: 10.1523/JNEUROSCI.3066-09.2010

Izawa, J. & Shadmehr, R. (2011). Learning from sensory and reward prediction errors during motor adaptation. *PLoS Computational Biology*, *7*(3), e1002012. doi: 10.1371/journal.pcbi.1002012

Jenkinson, M., Bannister, P., Brady, M. & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. doi: 10.1006/nimg.2002.1132

Jenkinson, M. & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156. doi: 10.1016/S1361-8415(01)00036-6

Jirenhed, D.-A., Bengtsson, F. & Hesslow, G. (2007). Acquisition, extinction, and reacquisition of a cerebellar cortical memory trace. *Journal of Neuroscience*, *27*(10), 2493–2502. doi: 10.1523/JNEUROSCI.4202-06.2007

Jordan, K., Heinze, H.-J., Lutz, K., Kanowski, M. & Jäncke, L. (2001). Cortical activations during the mental rotation of different visual objects. *NeuroImage*, *13*(1), 143–152. doi: 10.1006/nimg.2000.0677

Jordan, M. I. & Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. , *16*(3), 307–354. Retrieved from `http://doi.wiley.com/10.1207/s15516709cog1603_1` doi: 10.1207/s15516709cog1603_1

Joshua, M. & Lisberger, S. (2015). A tale of two species: Neural integration in zebrafish and monkeys. *Neuroscience*, *26*, 80-91. doi: 10.1016/j.neuroscience.2014.04.048

Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–292. doi: 10.2307/1914185

Kapogiannis, D., Campion, P., Grafman, J. & Wassermann, E. M. (2008). Reward-related activity in the human motor cortex. *European Journal of Neuroscience*, *27*(7), 1836–1842. doi: 10.1111/j.1460-9568.2008.06147.x

Kasuga, S., Telgen, S., Ushiba, J., Nozaki, D. & Diedrichsen, J. (2015). Learning feedback and feedforward control in a mirror-reversed visual environment. *Journal of Neurophysiology*, *114*(4), 2187–2193. doi: 10.1152/jn.00096.2015

Kawato, M. & Gomi, H. (1992). A computational model of four regions of the cerebellum based on feedback-error learning. , *68*(2), 95–103. doi: 10.1007/BF00201431

Keisler, A. & Shadmehr, R. (2010). A shared resource between declarative memory and motor memory. *Journal of Neuroscience*, *30*(44), 14817–14823. doi: 10.1523/JNEUROSCI.4160-10.2010

Kim, H. E., Parvin, D. E. & Ivry, R. B. (2019). The influence of task outcome on implicit motor learning. *eLife*, 28. doi: 10.7554/eLife.39882

Kim, S., Oh, Y. & Schweighofer, N. (2015). Between-trial forgetting due to interference and time in motor adaptation. *PLOS ONE*, *10*(11), e0142963. doi: 10.1371/journal.pone.0142963

Kitago, T., Ryan, S. L., Mazzoni, P., Krakauer, J. W. & Haith, A. M. (2013). Unlearning versus savings in visuomotor adaptation: comparing effects of washout, passage of time, and removal of errors on motor memory. *Frontiers in Human Neuroscience*, *7*. doi: 10.3389/fnhum.2013.00307

Klein-Flugge, M. C., Kennerley, S. W., Friston, K. & Bestmann, S. (2016). Neural signatures of value comparison in human cingulate cortex during decisions requiring an effort-reward trade-off. *Journal of Neuroscience*, *36*(39), 10002–10015. doi: 10.1523/JNEUROSCI.0292-16.2016

Kojima, Y. & Soetedjo, R. (2017). Selective reward affects the rate of saccade adaptation. *Neuroscience*, *355*, 113–125. doi: 10.1016/j.neuroscience.2017.04.048

Krakauer, J. W. (2009). Motor learning and consolidation: The case of visuomotor rotation. In D. Sternad (Ed.), *Progress in motor control* (Vol. 629, p. 405–421). Springer US.

Lago-Rodriguez, A. & Miall, R. C. (2016). Online visual feedback during error-free channel trials leads to active unlearning of movement dynamics: Evidence for adaptation to trajectory prediction errors. *Frontiers in Human Neuroscience*, *10*. doi: 10.3389/fnhum.2016.00472

Lakens, D., Adolfi, F. G., Albers, C. J., Anvari, F., Apps, M. A. J., Argamon, S. E., . . . et al. (2018). Justify your alpha. *Nature Human Behaviour*, *2*(3), 168–171. doi: 10.1038/s41562-018-0311-x

Leow, L.-A., de Rugy, A., Marinovic, W., Riek, S. & Carroll, T. J. (2016). Savings for visuomotor adaptation require prior history of error, not prior repetition of successful actions. *Journal of Neurophysiology*, *116*(4), 1603–1614. doi: 10.1152/jn.01055.2015

Leow, L.-A., Gunn, R., Marinovic, W. & Carroll, T. J. (2017). Estimating the implicit component of visuomotor rotation learning by constraining movement preparation time. *Journal of Neurophysiology*, jn.00834.2016. doi: 10.1152/jn.00834.2016

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

Lev-Ran, S., Shamay-Tsoory, S., Zangen, A. & Levkovitz, Y. (2012). Transcranial magnetic stimulation of the ventromedial prefrontal cortex impairs theory of mind learning. *European Psychiatry*, *27*(4), 285–289. doi: 10.1016/j.eurpsy.2010.11.008

Lewandowsky, S. & Farrell, S. B. (2011). Considering the data: What level of analysis? In *Computational modeling in cognition: Principles and practice* (p. 96-108). Sage Publications.

Liao, C.-M. & Masters, R. S. (2001). Analogy learning: A means to implicit motor learning. *Journal of Sports Sciences*, *19*(5), 307–319. doi: 10.1080/02640410152006081

Llewellyn, M. E., Thompson, K. R., Deisseroth, K. & Delp, S. L. (2010). Orderly recruitment of motor units under optical control in vivo. *Nature Medicine*, *16*(10), 1161–1165. doi: 10.1038/nm.2228

Loonis, R. F., Brincat, S. L., Antzoulatos, E. G. & Miller, E. K. (2017). A meta-analysis suggests different neural correlates for implicit and explicit learning. *Neuron*, *96*(2), 521-534.e7. doi: 10.1016/j.neuron.2017.09.032

Malfait, N. (2004). Is interlimb transfer of force-field adaptation a cognitive response to the sudden introduction of load? *Journal of Neuroscience*, *24*(37), 8084–8089. doi: 10.1523/JNEUROSCI.1742-04.2004

Manley, H., Dayan, P. & Diedrichsen, J. (2014). When money is not enough: Awareness, success, and variability in motor learning. *PLoS ONE*, *9*(1), e86580. doi: 10.1371/journal.pone.0086580

Manohar, S. G., Chong, T. T.-J., Apps, M. A., Batla, A., Stamelou, M., Jarman, P. R., ... Husain, M. (2015). Reward pays the cost of noise reduction in motor and cognitive control. *Current Biology*, *25*(13), 1707–1716. doi: 10.1016/j.cub.2015.05.038

Manohar, S. G., Muhammed, K., Fallon, S. J. & Husain, M. (2019). Motivation dynamically increases noise resistance by internal feedback during movement. *Neuropsychologia*, *123*, 19–29. doi: 10.1016/j.neuropsychologia.2018.07.011

Maris, E. & Oostenveld, R. (2007). Nonparametric statistical testing of eeg- and meg-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. doi: 10.1016/j.jneumeth.2007.03.024

Martin, T. A., Keating, J. G., Goodkin, H. P., Bastian, A. J. & Thach, W. T. (1996). Throwing while looking through prisms: I. focal olivocerebellar lesions impair adaptation. *Brain*, *119*(4), 1183–1198.

Martin, V., Scholz, J. P. & Schöner, G. (2009). Redundancy, self-motion, and motor control. *Neural Computation*, *21*(5), 1371–1414. doi: 10.1162/neco.2008.01-08-698

Matsunaga, K., Maruyama, A., Fujiwara, T., Nakanishi, R., Tsuji, S. & Rothwell, J. C. (2005). Increased corticospinal excitability after 5 hz rtms over the human supplementary motor area. *The Journal of Physiology*, *562*(1), 295–306. doi: 10.1113/jphysiol.2004.070755

Maunsell, J. H. (2004). Neuronal representations of cognitive state: reward or attention? *Trends in Cognitive Sciences*, *8*(6), 261–265. doi: 10.1016/j.tics.2004.04.003

Mawase, F., Uehara, S., Bastian, A. J. & Celnik, P. (2017). Motor learning enhances use-dependent plasticity. *The Journal of Neuroscience*, *37*(10), 2673–2685. doi: 10.1523/JNEUROSCI.3303-16.2017

Mawase, F., Wymbs, N., Uehara, S. & Celnik, P. (2016). Reward gain model describes cortical use-dependent plasticity. In *Engineering in medicine and biology society (embc), 2016 ieee 38th annual international conference of the* (p. 5–8). IEEE.

Maxwell, J., Masters, R. & Eves, F. (2000). From novice to no know-how: A longitudinal study of implicit motor learning. *Journal of Sports Sciences*, *18*(2), 111–120. doi: 10.1080/026404100365180

Maxwell, J., Masters, R., Kerr, E. & Weedon, E. (2001). The implicit benefit of learning without errors. *The Quarterly Journal of Experimental Psychology Section A*, *54*(4), 1049–1068. doi: 10.1080/713756014

Mazzoni, P., Hristova, A. & Krakauer, J. W. (2007). Why don't we move faster? parkinson's disease, movement vigor, and implicit motivation. *Journal of Neuroscience*, *27*(27), 7105–7116. doi: 10.1523/JNEUROSCI.0264-07.2007

Mazzoni, P. & Krakauer, J. W. (2006). An implicit plan overrides an explicit strategy during visuomotor adaptation. *Journal of Neuroscience*, *26*(14), 3642–3645. doi: 10.1523/JNEUROSCI.5317-05.2006

McDougle, S. D., Bond, K. M. & Taylor, J. A. (2015). Explicit and implicit processes constitute the fast and slow processes of sensorimotor learning. *Journal of Neuroscience*, *35*(26), 9568–9579. doi: 10.1523/JNEUROSCI.5061-14.2015

McDougle, S. D. & Taylor, J. A. (2019). Dissociable cognitive strategies for sensorimotor learning. *Nature Communications*, *10*(1). doi: 10.1038/s41467-018-07941-0

McNab, F. & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience*, *11*(1), 103–107. doi: 10.1038/nn2024

Miall, R. C., Christensen, L. O., Cain, O. & Stanley, J. (2007). Disruption of state estimation in the human lateral cerebellum. *PLoS biology*, *5*(11), e316.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P. & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? a latent-variable analysis. *Journal of Experimental Psychology: General*, *130*(4), 621-640. doi: 10.1037/0096-3445.130.4.621

Modchalingam, S., Vachon, C. M., 't Hart, B. M. & Henriques, D. Y. P. (2019). The effects of awareness of the perturbation during motor adaptation on hand localization. , *14*(8), 20. doi: 10.1371/journal.pone.0220884

Morehead, J. R., Qasim, S. E., Crossley, M. J. & Ivry, R. (2015). Savings upon re-aiming in visuomotor adaptation. *Journal of Neuroscience*, *35*(42), 14386–14396. doi: 10.1523/JNEUROSCI.1046-15.2015

Morehead, J. R., Taylor, J. A., Parvin, D. & Ivry, R. B. (2017). Characteristics of implicit sensorimotor adaptation revealed by task-irrelevant clamped feedback. *Journal of Cognitive Neuroscience*, 1–14. doi: 10.1162/jocn_a_01108

Mulavara, A. P., Feiveson, A. H., Fiedler, J., Cohen, H., Peters, B. T., Miller, C., ... Bloomberg, J. J. (2010). Locomotor function after long-duration space flight: effects and motor learning during recovery. *Experimental Brain Research*, *202*(3), 649–659. doi: 10.1007/s00221-010-2171-0

Mussa-Ivaldi, F., Hogan, N. & Bizzi, E. (1985). Neural, mechanical, and geometric factors subserving arm posture in humans. *The Journal of Neuroscience*, *5*(10), 2732–2743. doi: 10.1523/JNEUROSCI.05-10-02732.1985

Nakahara, H. & Hikosaka, O. (2012). Learning to represent reward structure: A key to adapting to complex environments. *Neuroscience Research*, *74*(3–4), 177–183. doi: 10.1016/j.neures.2012.09.007

Nichols, T. E. & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, *15*(1), 1–25. doi: 10.1002/hbm.1058

Omrani, M., Murnaghan, C. D., Pruszynski, J. A. & Scott, S. H. (2016). Distributed task-specific processing of somatosensory feedback for voluntary motor control. *eLife*, *5*. doi: 10.7554/eLife.13141

Open Science Collaboration, O. S. C. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716–aac4716. doi: 10.1126/science.aac4716

Orban de Xivry, J.-J. & Lefèvre, P. (2015). Formation of model-free motor memories during motor adaptation depends on perturbation schedule. *Journal of Neurophysiology*, *113*(7), 2733–2741. doi: 10.1152/jn.00673.2014

Orban de Xivry, J.-J., Legrain, V. & Lefèvre, P. (2017). Overlap of movement planning and movement execution reduces reaction time. *Journal of Neurophysiology*, *117*(1), 117–122. doi: 10.1152/jn.00728.2016

Otto, A. R., Gershman, S. J., Markman, A. B. & Daw, N. D. (2013). The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychological science*, *24*(5), 751–761.

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A. & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences*, *110*(52), 20941–20946. doi: 10.1073/pnas.1312011110

Otto, A. R., Skatova, A., Madlon-Kay, S. & Daw, N. D. (2015). Cognitive control predicts use of model-based reinforcement learning. *Journal of Cognitive Neuroscience*, *27*(2), 319–333. doi: 10.1162/jocn_a_00709

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

Pasquereau, B., Nadjar, A., Arkadir, D., Bezard, E., Goillandeau, M., Bioulac, B., ... Boraud, T. (2007). Shaping of motor responses by incentive values through the basal ganglia. *Journal of Neuroscience*, *27*(5), 1176–1183. doi: 10.1523/JNEUROSCI.3745-06.2007

Pearson-Fuhrhop, K. M., Minton, B., Acevedo, D., Shahbaba, B. & Cramer, S. C. (2013). Genetic variation in the human brain dopamine system influences motor learning and its modulation by l-dopa. *PLoS ONE*, *8*(4), e61197. doi: 10.1371/journal.pone.0061197

Pekny, S. E., Criscimagna-Hemminger, S. E. & Shadmehr, R. (2011). Protection and expression of human motor memories. *Journal of Neuroscience*, *31*(39), 13829–13839. doi: 10.1523/JNEUROSCI.1704-11.2011

Pekny, S. E., Izawa, J. & Shadmehr, R. (2015). Reward-dependent modulation of movement variability. *Journal of Neuroscience*, *35*(9), 4015–4024. doi: 10.1523/JNEUROSCI.3244-14.2015

Perreault, E. J., Kirsch, R. F. & Crago, P. E. (2002). Voluntary control of static endpoint stiffness during force regulation tasks. *J. Neurophysiol*, *87*(6), 2808–2816. doi: 10.1152/jn.00590.2001

Peters, M. & Battista, C. (2008). Applications of mental rotation figures of the shepard and metzler type and description of a mental rotation stimulus library. *Brain and Cognition*, *66*(3), 260–264. doi: 10.1016/j.bandc.2007.09.003

Pinto, L., Goard, M. J., Estandian, D., Xu, M., Kwan, A. C., Lee, S.-H., … Dan, Y. (2013). Fast modulation of visual perception by basal forebrain cholinergic neurons. *Nature Neuroscience*, *16*(12), 1857–1863. doi: 10.1038/nn.3552

Porter, J. D., Baker, R. S., Ragusa, R. J. & Brueckner, J. K. (1995). Extraocular muscles: Basic and clinical aspects of structure and function. *Survey of Ophthalmology*, *39*(6), 451–484. doi: 10.1016/S0039-6257(05)80055-4

Pruszynski, J. A., Kurtzer, I., Nashed, J. Y., Omrani, M., Brouwer, B. & Scott, S. H. (2011). Primary motor cortex underlies multi-joint integration for fast feedback control. *Nature*, *478*(7369), 387–390. doi: 10.1038/nature10436

Quattrocchi, G., Greenwood, R., Rothwell, J. C., Galea, J. M. & Bestmann, S. (2017). Reward and punishment enhance motor adaptation in stroke. *Journal of Neurology, Neurosurgery & Psychiatry*, *88*(9), 730–736. doi: 10.1136/jnnp-2016-314728

Quattrocchi, G., Monaco, J., Ho, A., Irmen, F., Strube, W., Ruge, D., … Galea, J. M. (2018). Pharmacological dopamine manipulation does not alter reward-based improvements in memory retention during a visuomotor adaptation task. *eneuro*, *5*(3), ENEURO.0453-17.2018. doi: 10.1523/ENEURO.0453-17.2018

Quist, J. F., Barr, C. L., Schachar, R., Roberts, W., Malone, M., Tannock, R., … Kennedy, J. L. (2003). The serotonin 5-ht1b receptor gene and attention deficit hyperactivity disorder. *Molecular Psychiatry*, *8*(1), 98–102. doi: 10.1038/sj.mp.4001244

Ramkumar, P., Dekleva, B., Cooler, S., Miller, L. & Kording, K. (2016). Premotor and motor cortices encode reward. *PLOS ONE*, *11*(8), e0160851. doi: 10.1371/journal.pone .0160851

Rascol, O., Sabatini, U., Chollet, F., Fabre, N., Senard, J. M., Montastruc, J. L., ... Rascol, A. (1994). Normal activation of the supplementary motor area in patients with parkinson's disease undergoing long-term treatment with levodopa. *Journal of Neurology, Neurosurgery & Psychiatry*, *57*(5), 567–571. doi: 10.1136/jnnp.57.5.567

Reis, J., Schambra, H. M., Cohen, L. G., Buch, E. R., Fritsch, B., Zarahn, E., ... Krakauer, J. W. (2009). Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation. *Proceedings of the National Academy of Sciences*, *106*(5), 1590–1595. doi: 10.1073/pnas.0805413106

Reppert, T. R., Lempert, K. M., Glimcher, P. W. & Shadmehr, R. (2015). Modulation of saccade vigor during value-based decision making. *Journal of Neuroscience*, *35*(46), 15369-15378. doi: 10.1523/JNEUROSCI.2621-15.2015

Reppert, T. R., Rigas, I., Herzfeld, D. J., Sedaghat-Nejad, E., Komogortsev, O. & Shadmehr, R. (2018). Movement vigor as a traitlike attribute of individuality. *Journal of Neurophysiology*, *120*(2), 741–757. doi: 10.1152/jn.00033.2018

Robertson, E. M. (2007). The serial reaction time task: Implicit motor skill learning? *Journal of Neuroscience*, *27*(38), 10073–10075. doi: 10.1523/JNEUROSCI.2747-07.2007

Robinson, D. A. (1964). The mechanics of human saccadic eye movement. *The Journal of Physiology*, *174*(2), 245-264. doi: 10.1113/jphysiol.1964.sp007485

Roesch, M. R. & Olson, C. R. (2003). Impact of expected reward on neuronal activity in prefrontal cortex, frontal and supplementary eye fields and premotor cortex. *Journal of Neurophysiology*, *90*(3), 1766–1789. doi: 10.1152/jn.00019.2003

Roesch, M. R. & Olson, C. R. (2004). Neuronal activity related to reward value and motivation in primate frontal cortex. *Science*, *304*(5668), 307–310.

Rorden, C. & Brett, M. (2000). Stereotaxic display of brain lesions. *Behavioural Neurology*, *12*(4), 191–200. doi: 10.1155/2000/421719

Rousselet, G. A., Foxe, J. J. & Bolam, J. P. (2016). A few simple steps to improve the description of group results in neuroscience. *European Journal of Neuroscience*, *44*(9), 2647–2651. doi: 10.1111/ejn.13400

Rousselet, G. A., Pernet, C. R. & Wilcox, R. R. (2017). Beyond differences in means: robust graphical methods to compare two groups in neuroscience. *European Journal of Neuroscience*, *46*(2), 1738–1748. doi: 10.1111/ejn.13610

Ruitenberg, M. F. L., Koppelmans, V., De Dios, Y. E., Gadd, N. E., Wood, S. J., Reuter-Lorenz, P. A., … Seidler, R. D. (2018). Neural correlates of multi-day learning and savings in sensorimotor adaptation. *Scientific Reports*, *8*(1), 14286. doi: 10.1038/s41598-018-32689-4

Saijo, N. & Gomi, H. (2010). Multiple motor learning strategies in visuomotor rotation. *PLoS ONE*, *5*(2), e9399. doi: 10.1371/journal.pone.0009399

Sarter, M., Gehring, W. J. & Kozak, R. (2006). More attention must be paid: The neurobiology of attentional effort. *Brain Research Reviews*, *51*(2), 145–160. doi: 10.1016/j.brainresrev.2005.11.002

Sarter, M., Hasselmo, M. E., Bruno, J. P. & Givens, B. (2005). Unraveling the attentional functions of cortical cholinergic inputs: interactions between signal-driven and cognitive modulation of signal detection. *Brain Research Reviews*, *48*(1), 98–111. doi: 10.1016/j.brainresrev.2004.08.006

Sawaki, R., Luck, S. J. & Raymond, J. E. (2015). How attention changes in response to incentives. *Journal of Cognitive Neuroscience*, *27*(11), 2229–2239. doi: 10.1162/jocn_a_00847

Scheidt, R. A., Conditt, M. A., Secco, E. L. & Mussa-Ivaldi, F. A. (2005). Interaction of visual and proprioceptive feedback during adaptation of human reaching movements. *Journal of Neurophysiology*, *93*(6), 3200–3213. doi: 10.1152/jn.00947.2004

Schielzeth, H. & Nakagawa, S. (2013). Nested by design: model fitting and interpretation in a mixed model era. *Methods in Ecology and Evolution*, *4*(1), 14-24. doi: 10.1111/j.2041-210x.2012.00251.x

Selen, L. P. J., Franklin, D. W. & Wolpert, D. M. (2009). Impedance control reduces instability that arises from motor noise. *Journal of Neuroscience*, *29*(40), 12606–12616. doi: 10.1523/JNEUROSCI.2826-09.2009

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

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

Shepard, N., Roger & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703. doi: 10.1126/science.171.3972.701

Shirota, Y., Hamada, M., Terao, Y., Ohminami, S., Tsutsumi, R., Ugawa, Y. & Hanajima, R. (2012). Increased primary motor cortical excitability by a single-pulse transcranial magnetic stimulation over the supplementary motor area. *Experimental Brain Research*, *219*(3), 339–349. doi: 10.1007/s00221-012-3095-7

Shmuelof, L., Huang, V. S., Haith, A. M., Delnicki, R. J., Mazzoni, P. & Krakauer, J. W. (2012). Overcoming motor "forgetting" through reinforcement of learned actions. *Journal of Neuroscience*, *32*(42), 14617–14621a. doi: 10.1523/JNEUROSCI.2184-12.2012

Shmuelof, L., Yang, J., Caffo, B., Mazzoni, P. & Krakauer, J. W. (2014). The neural correlates of learned motor acuity. *Journal of Neurophysiology*, *112*(4), 971–980. doi: 10.1152/jn.00897.2013

Sidarta, A., van Vugt, F. & Ostry, D. J. (2018). Somatosensory working memory in human reinforcement-based motor learning. *Journal of Neurophysiology*, 41. doi: 10.1152/jn.00442.2018

Simon, D. A. & Daw, N. D. (2011). Neural correlates of forward planning in a spatial decision task in humans. *Journal of Neuroscience*, *31*(14), 5526–5539. doi: 10.1523/JNEUROSCI.4647-10.2011

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

Sobel, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. *Sociological Methodology*, *16*, 159. doi: 10.2307/270922

Sohn, J.-w. & Lee, D. (2006). Effects of reward expectancy on sequential eye movements in monkeys. *Neural Networks*, *19*(8), 1181–1191. doi: 10.1016/j.neunet.2006.04.005

Song, Y. & Smiley-Oyen, A. L. (2017). Probability differently modulating the effects of reward and punishment on visuomotor adaptation. *Experimental Brain Research*, *235*(12), 3605–3618. doi: 10.1007/s00221-017-5082-5

Sosnik, R., Hauptmann, B., Karni, A. & Flash, T. (2004). When practice leads to co-articulation: the evolution of geometrically defined movement primitives. *Experimental Brain Research*, *156*(4), 422–438. doi: 10.1007/s00221-003-1799-4

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, *82*(3), 171–177. doi: 10.1016/j.nlm.2004.06.005

Stanford, A. D., Luber, B., Unger, L., Cycowicz, Y. M., Malaspina, D. & Lisanby, S. H. (2013). Single pulse tms differentially modulates reward behavior. *Neuropsychologia*, *51*(14), 3041–3047. doi: 10.1016/j.neuropsychologia.2013.09.016

Stanley, J. & Krakauer, J. W. (2013). Motor skill depends on knowledge of facts. *Frontiers in Human Neuroscience*, *7*. doi: 10.3389/fnhum.2013.00503

Suchan, B., Botko, R., Gizewski, E., Forsting, M. & Daum, I. (2006). Neural substrates of manipulation in visuospatial working memory. *Neuroscience*, *139*(1), 351–357. doi: 10.1016/j.neuroscience.2005.08.020

Summerside, E. M., Shadmehr, R. & Ahmed, A. A. (2018). Vigor of reaching movements: reward discounts the cost of effort. *Journal of Neurophysiology*, *119*(6), 2347–2357. doi: 10.1152/jn.00872.2017

Sun, R., Slusarz, P. & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychological Review*, *112*(1), 159–192. doi: 10.1037/0033-295X.112.1.159

Sutton, R. S. (1990). Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In *In proceedings of the seventh international conference on machine learning* (p. 216–224). Morgan Kaufmann.

Sutton, R. S. & Barto, A. (1998). *Reinforcement learning: An introduction.* A Bradford Book.

Sutton, R. S., Szepesvári, C., Geramifard, A. & Bowling, M. (2008). Dyna-style planning with linear function approximation and prioritized sweeping. In *Proceedings of the 24th conference on uncertainty in artificial intelligence.*

Takikawa, Y., Kawagoe, R., Itoh, H., Nakahara, H. & Hikosaka, O. (2002). Modulation of saccadic eye movements by predicted reward outcome. *Experimental Brain Research*, *142*(2), 284–291. doi: 10.1007/s00221-001-0928-1

Taylor, J. A. & Ivry, R. B. (2011). Flexible cognitive strategies during motor learning. *PLoS Computational Biology*, *7*(3), e1001096. doi: 10.1371/journal.pcbi.1001096

Taylor, J. A. & Ivry, R. B. (2014). Cerebellar and prefrontal cortex contributions to adaptation, strategies, and reinforcement learning. In *Progress in brain research* (Vol. 210, p. 217–253). Elsevier.

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

Telgen, S., Parvin, D. & Diedrichsen, J. (2014). Mirror reversal and visual rotation are learned and consolidated via separate mechanisms: Recalibrating or learning de novo? *Journal of Neuroscience*, *34*(41), 13768–13779. doi: 10.1523/JNEUROSCI.5306-13.2014

Thabit, M. N., Nakatsuka, M., Koganemaru, S., Fawi, G., Fukuyama, H. & Mima, T. (2011). Momentary reward induce changes in excitability of primary motor cortex. *Clinical Neurophysiology*, *122*(9), 1764–1770. doi: 10.1016/j.clinph.2011.02.021

Therrien, A. S., Wolpert, D. M. & Bastian, A. J. (2016). Effective reinforcement learning following cerebellar damage requires a balance between exploration and motor noise. *Brain*, *139*(1), 101-114. doi: 10.1093/brain/awv329

Therrien, A. S., Wolpert, D. M. & Bastian, A. J. (2018). Increasing motor noise impairs reinforcement learning in healthy individuals. *eneuro*, *5*(3), ENEURO.0050-18.2018. doi: 10.1523/ENEURO.0050-18.2018

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

Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature Neuroscience*, *7*(9), 907–915. doi: 10.1038/nn1309

Todorov, E. (2005). Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system. *Neural Computation*, *17*(5), 1084–1108. doi: 10.1162/0899766053491887

Tosoni, A., Shulman, G. L., Pope, A. L., McAvoy, M. P. & Corbetta, M. (2013). Distinct representations for shifts of spatial attention and changes of reward contingencies in the human brain. *Cortex*, *49*(6), 1733–1749. doi: 10.1016/j.cortex.2012.03.022

Tricomi, E., Balleine, B. W. & O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *European Journal of Neuroscience*, *29*(11), 2225–2232. doi: 10.1111/j.1460-9568.2009.06796.x

Tseng, Y.-w., Diedrichsen, J., Krakauer, J. W., Shadmehr, R. & Bastian, A. J. (2007). Sensory prediction errors drive cerebellum-dependent adaptation of reaching. *Journal of Neurophysiology*, *98*(1), 54–62. doi: 10.1152/jn.00266.2007

Ueyama, Y. & Miyashita, E. (2013). Signal-dependent noise induces muscle co-contraction to achieve required movement accuracy: A simulation study with an optimal control. *Current Bioinformatics*, *8*(1), 16–24. doi: 10.2174/1574893611308010005

Ueyama, Y. & Miyashita, E. (2014). Optimal feedback control for predicting dynamic stiffness during arm movement. *IEEE Transactions on Industrial Electronics*, *61*(2), 1044–1052. doi: 10.1109/TIE.2013.2273473

Ueyama, Y., Miyashita, E., Pham, T. D., Zhou, X., Tanaka, H., Oyama-Higa, M., ... Jia, X. (2011). Cocontraction of pairs of muscles around joints may improve an accuracy of a reaching movement: a numerical simulation study. In (p. 73–82). doi: 10.1063/1.3596629

van Beers, R. J., Haggard, P. & Wolpert, D. M. (2004). The role of execution noise in movement variability. *Journal of Neurophysiology*, *91*(2), 1050-1063. doi: 10.1152/jn.00652.2003 van der Kooij, K., Oostwoud Wijdenes, L., Rigterink, T., Overvliet, K. E. & Smeets, J. B. J. (2018). Reward abundance interferes with error-based learning in a visuomotor adaptation task. *PLOS ONE*, *13*(3), e0193002. doi: 10.1371/journal.pone.0193002

van der Kooij, K. & Overvliet, K. E. (2016). Rewarding imperfect motor performance reduces adaptive changes. *Experimental Brain Research*, *234*(6), 1441–1450. doi: 10.1007/s00221-015-4540-1

van der Kooij, K., van Dijsseldonk, R., van Veen, M., Steenbrink, F., de Weerd, C. & Overvliet, K. E. (2019). Gamification as a sustainable source of enjoyment during balance and gait exercises. *Frontiers in Psychology*, *10*. doi: 10.3389/fpsyg.2019.00294

Van Gisbergen, J. A., Robinson, D. A. & Gielen, S. (1981). A quantitative analysis of generation of saccadic eye movements by burst neurons. *Journal of Neurophysiology*, *45*(3), 417–442. doi: 10.1152/jn.1981.45.3.417

Vilis, T. & Hore, J. (1980). Central neural mechanisms contributing to cerebellar tremor produced by limb perturbations. *Journal of Neurophysiology*, *43*(2), 279–291.

Vindras, P., Desmurget, M., Prablanc, C. & Viviani, P. (1998). Pointing errors reflect biases in the perception of the initial hand position. *Journal of neurophysiology*, *79*(6), 3290–3294.

Wachter, T., Lungu, O. V., Liu, T., Willingham, D. T. & Ashe, J. (2009). Differential effect of reward and punishment on procedural learning. *Journal of Neuroscience*, *29*(2), 436–443. doi: 10.1523/JNEUROSCI.4132-08.2009

Walker, M. P. & van der Helm, E. (2009). Overnight therapy? the role of sleep in emotional brain processing. *Psychological Bulletin*, *135*(5), 731–748. doi: 10.1037/a0016570

Weiler, J., Gribble, P. L. & Pruszynski, J. A. (2019). Spinal stretch reflexes support efficient hand control. *Nature Neuroscience*, *22*(4), 529–533. doi: 10.1038/s41593-019-0336-0

Werner, S., van Aken, B. C., Hulst, T., Frens, M. A., van der Geest, J. N., Strüder, H. K. & Donchin, O. (2015). Awareness of sensorimotor adaptation to visual rotations of different size. *PLOS ONE*, *10*(4), e0123321. doi: 10.1371/journal.pone.0123321

Wolpert, D. M. & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural networks*, *11*(7), 1317–1329.

Wolpert, D. M. & Miall, R. C. (1996). Forward models for physiological motor control. *Neural Networks: The Official Journal of the International Neural Network Society*, *9*(8), 1265–1279.

Wolpert, D. M., Miall, R. C. & Kawato, M. (1998). Internal models in the cerebellum. *Trends in cognitive sciences*, *2*(9), 338–347.

Wong, A. L., Marvel, C. L., Taylor, J. A. & Krakauer, J. W. (2019). Can patients with cerebellar disease switch learning mechanisms to reduce their adaptation deficits. *Brain*, *142*(1), 662–673. doi: 10.1093/brain/awy334

Wunderlich, K., Dayan, P. & Dolan, R. J. (2012). Mapping value based planning and extensively trained choice in the human brain. *Nature Neuroscience*, *15*(5), 786–791. doi: 10.1038/nn.3068

Xu-Wilson, M., Zee, D. S. & Shadmehr, R. (2009). The intrinsic value of visual information affects saccade velocities. *Experimental Brain Research*, *196*(4), 475–481. doi: 10.1007/s00221-009-1879-1

Yang, Y. & Lisberger, S. G. (2014). Role of plasticity at different sites across the time course of cerebellar motor learning. *Journal of Neuroscience*, *34*(21), 7077–7090. doi: 10.1523/JNEUROSCI.0017-14.2014

Zenon, A., Sidibe, M. & Olivier, E. (2015). Disrupting the supplementary motor area makes physical effort appear less effortful. *Journal of Neuroscience*, *35*(23), 8737–8744. doi: 10.1523/JNEUROSCI.3789-14.2015

Zhao, Y., Hessburg, J. P., Asok Kumar, J. N. & Francis, J. T. T. (2018). Paradigm shift in sensorimotor control research and brain machine interface control: The influence of context on sensorimotor representations. *Frontiers in Neuroscience*. doi: 10.1101/239814

Zuur, A. F. (2009). *Mixed effects models and extensions in ecology with r*. Springer.

Zuur, A. F., Ieno, E. N. & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems: Data exploration. *Methods in Ecology and Evolution*, *1*(1), 3-14. Retrieved from `http://doi.wiley.com/10.1111/j.2041-210X.2009.00001.x` doi: 10.1111/j.2041-210X.2009.00001.x