ACTION SELECTION IN MOTOR CONTROL: ERROR, REWARD, AND TIME

by
Pavan Anil Vaswani

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
September, 2014

Abstract

Each action our bodies execute is the consequence of a complex process of decision making by the brain. At each moment, a multitude of actions are available to the brain to place the body in a more desirable state. How, then, does the brain decide what movement to make, and why? In general, we presented the motor system with tasks where several actions could be used to successfully accomplish a task, and observed what solution healthy people and people with movement disorders adopted. Our goal was to understand the costs and tradeoffs involved in the selection of movement. While many influences likely contribute, we focused on the role of three factors: errors, reward, and time. We found that errors played a dominant role in the selection of action. First, we used error-clamp trials, where we injected artificial redundancy and controlled the errors and rewards produced by each action, to examine the effect of the error on the selection of actions. We found that the policy underlying the selection of movement changes when the brain detects a change in the distribution of errors in the task. People typically learned from sensory prediction error, but, in the presence of artificial redundancy, could use other policies. Next, we created and examined situations where two factors were opposed, to study the relationship between them. When we pit the error-based and reward-based learning systems against each other, we found that the former dominated behavior. When we used a paradigm where time and reward were

opposed, we found interindividual differences in the way people discounted reward in time: people who were willing to wait longer to receive a greater probability of success also made slower saccadic eye movements, suggesting a common cost of time in decision making in motor control. Finally, we found that a tradeoff between endpoint error and time could account for the way healthy people and people with the movement disorder ataxia-telangiectasia made eye movements. Overall, our results demonstrate that the brain predominantly seeks to reduce error but also works to achieve success and move quickly. It is the balance of these sometimes-conflicting desires that guides the precision and elegance of human movement.

Thesis Advisor: Dr. Reza Shadmehr *Secondary Reader*: Dr. John Krakauer

Thesis Committee: Dr. John Krakauer, Dr. Amy Bastian, Dr. Thomas Crawford

Acknowledgements

I am extremely fortunate to have had the support of many friends and mentors. I am grateful to my advisor, Dr. Reza Shadmehr, for his mentorship, dedicated pursuit of knowledge, and open door to discussion. He made sure I was always thinking and immersing myself in the pursuit of truth. I am grateful to Dr. John Krakauer for his always vigorous discussion and collaboration. Dr. Tom Crawford has been a wonderful advisor, friend, and collaborator. Dr. Amy Bastian's insightful comments have greatly advanced this work.

The members of the Laboratory for Computational Motor Control have been great friends and scientific colleagues. Coffee-fueled and chalk-filled conversations with David Herzfeld have made day-to-day research invigorating and exciting. I appreciate the camaraderie of Thomas Reppert, Sarah Pekny, Scott Albert, Mollie Marko, Dr. Yousef Salimpour, Jennie Choi, Dr. Sarah Hemminger, and Dr. Adrian Haith. Dr. Robert Siliciano, Dr. Andrea Cox, Sharon Welling, Martha Buntin, and Bernadine Harper of the MD PhD program have also been wonderful supporters.

Jennifer Wright, Dr. Howard Lederman, Dr. Maureen Lefton-Greif, and Dr. Sharon McGrath-Morrow of the A-T Children's Center have been amazing collaborators and colleagues. Jennifer Thornton, Cynthia Rothblum, and the A-T

Children's Project provided financial support and went above and beyond to help with these experiments.

I cannot imagine having more amazing friends than Kathy Wei, Sam Burden, and Elias and Devin Le Brun-Cusack. Weddings, camping trips, closing out Pho restaurants, and email conversations have provided fond memories and inspiration.

I am indebted to Dr. Megan Rybarczyk for her patience, kindness, dedication, and support.

Finally, I am deeply grateful to the many volunteers and their families for their participation in the studies that form the basis for this dissertation. Their contribution has been invaluable.

Dedicated to my family

My mother

My first teacher, my first critic, and my greatest supporter

For her unconditional love, patience, confidence, wisdom, and dedication

And for teaching me how to think, and to always do my best

My father

My first hero and my first mentor

For his guidance, belief, inspiration, sincerity, and encouragement

And for teaching me the importance of hard work and self sufficiency

And

My brother

For being my first and oldest friend

Table of Contents

ABSTRACT ..... **II**

ACKNOWLEDGEMENTS ..... **IV**

CHAPTER 1 : INTRODUCTION **1**

SECTION 1.1 : REDUNDANCY IN THE SELECTION OF MOVEMENT 1

SECTION 1.2 : TOOLS TO DISAMBIGUATE GOALS IN MOTOR CONTROL 3

Perturbations *4*

Interindividual differences *5*

Movement disorders *7*

SECTION 1.3 : SCOPE OF THIS WORK 7

CHAPTER 2 : DECAY OF MOTOR MEMORIES IN THE ABSENCE OF ERROR **11**

SECTION 2.1 : MOTIVATION 11

SECTION 2.2 : METHODS 13

Experiment 1: Reaching with constant error-clamp *15*

Experiments 2: Shooting with constant error-clamp *16*

Experiments 3: Shooting with variable error-clamp *17*

Adaptation Index *18*

Decay of force in error-clamp trials *19*

Statistics *19*

SECTION 2.3 : RESULTS 20

Effect of history of training on rate and endpoint of decay 20

A lag to the start of the decay 24

Increasing the lag 28

Predicting an individual's lag 34

SECTION 2.4 : DISCUSSION 34

CHAPTER 3 : EXPLORATORY ESCAPE FROM PERSISTENT ERRORS 41

SECTION 3.1 : MOTIVATION 41

SECTION 3.2 : METHODS 43

Participants 43

Paradigm 43

State-space model 47

Data Analysis 48

Exploration 49

SECTION 3.3 : RESULTS 50

State-space model accounts for persistent residual error 52

Reversion to baseline incompletely predicted by a state-space model 53

Exploratory behavior in non-zero error clamps 56

Learning from error in error clamps with non-zero mean and non-zero variance ... 59

Quantifying exploration 64

SECTION 3.4 : DISCUSSION 66

Conclusion 72

CHAPTER 4 : MOTOR ADAPTATION WHEN MAXIMIZING REWARD IS IN CONFLICT WITH MINIMIZING ERRORS 73

SECTION 4.1 : MOTIVATION 73

SECTION 4.2 : METHODS 77

Perturbations 79

Experiment 1: Pretraining 81

Experiment 2: Repetition & reinforcement 82

Experiment 3: Reduced error feedback 83

Experiment 4: Reward in the absence of prediction error 84

Experiment 5: Sensory and reward prediction errors 86

Adaptation Index 87

Statistics 88

SECTION 4.3 : RESULTS 89

Experiment 1: Do subjects need to learn one solution first? 90

Experiment 2: Do subjects need to reinforce one solution? 94

Experiment 3: Can we make the error signal less salient? 97

Experiment 4: Does reward matter in the task? 100

Experiment 5: Can we modulate sensory and reward prediction errors? 105

Summary 109

SECTION 4.4 : DISCUSSION 111

CHAPTER 5 : VIGOR OF MOVEMENTS AND THE TEMPORAL DISCOUNTING OF REWARD 115

SECTION 5.1 : MOTIVATION 115

SECTION 5.2 : METHODS 117

Movement vigor 118

Temporal discount function 121

Modeling 126

SECTION 5.3 : RESULTS 126

Between-subject differences in movement velocities 127

Movement vigor 129

Estimating the temporal discount function 131

Relationship between vigor and willingness to wait 141

Individual differences in valuations of immediate reward 144

Reaction to the instruction-change cue 146

Psychological profile 148

SECTION 5.4 : DISCUSSION 149

Reaction time and vigor 153

Impulsivity and vigor 153

CHAPTER 6 : OPTIMAL CONTROL OF SACCADES IN ATAXIA-TELANGIECTASIA 155

SECTION 6.1 : MOTIVATION 155

SECTION 6.2 : METHODS 157

Paradigm *157*

Optimal Feedback Controller *160*

Simulations *162*

Statistics *163*

SECTION 6.3 : RESULTS 163

Saccade dynamics *166*

Signal dependent noise *170*

Optimal Feedback Controller *173*

Saccade gain in other movement disorders *177*

SECTION 6.4 : DISCUSSION 179

CHAPTER 7 : CONCLUSION 184

APPENDIX A: MAKING MULTIPLE SACCADES TO A TARGET: OPTIMAL FEEDBACK CONTROLLER 187

Problem statement *187*

Optimal Feedback Controller *189*

REFERENCES 196

CURRICULUM VITAE 206

List of Figures

FIGURE 2.1: EXPERIMENTAL PARADIGM 14

FIGURE 2.2: LEARNING AND DECAY IN REACHING MOVEMENTS 21

FIGURE 2.3: LAG TO DECAY IN REACHING MOVEMENTS 24

FIGURE 2.4: CHANGE IN BEHAVIOR 27

FIGURE 2.5: LAG TO DECAY IN SHOOTING MOVEMENTS 29

FIGURE 2.6: LAG TO DECAY WITH VARIABLE ERROR-CLAMP TRIALS 32

FIGURE 2.7: CUMULATIVE PROBABILITY DISTRIBUTION OF THE LAG TO DECAY 33

FIGURE 3.1**:** EXPERIMENTAL PARADIGM 45

FIGURE 3.2: BEHAVIOR WITH EXTENDED TRAINING AND IN CLAMP TRIALS TO ZERO 53

FIGURE 3.3: BEHAVIOR IN CLAMP TRIALS WITH NON-ZERO MEAN AND VARIABILITY 58

FIGURE 3.4: EXPLORATION IN CLAMP TRIALS 65

FIGURE 4.1: INTUITION 75

FIGURE 4.2: EXPERIMENTAL PARADIGM 80

FIGURE 4.3: EXPERIMENT 1: PRETRAINING 91

FIGURE 4.4: EXPERIMENT 2: ADDITIONAL REINFORCEMENT 96

FIGURE 4.5: EXPERIMENT 3: REDUCING ERROR 98

FIGURE 4.6: EXPERIMENT 4: THE ROLE OF REWARD IN THE ABSENCE OF SENSORY PREDICTION ERROR. 101

FIGURE 4.7: EXPERIMENT 5: SENSORY AND REWARD PREDICTION ERROR 107

FIGURE 4.8: SUMMARY OF BEHAVIOR 110

FIGURE 5.1: EXPERIMENTAL PROTOCOL 119

FIGURE 5.2: VIGOR OF SACCADES 128

FIGURE 5.3: DECISION MAKING OF SIMULATED SUBJECTS DURING EACH TRIAL 135

FIGURE 5.4: DECISION MAKING OF SIMULATED SUBJECTS DURING THE EXPERIMENT ... 139

FIGURE 5.5: WILLINGNESS TO WAIT AND MOVEMENT VIGOR 142

FIGURE 6.1: SACCADES OF A-T AND CONTROL SUBJECTS 159

FIGURE 6.2: INTERSACCADIC INTERVAL 165

FIGURE 6.3: SACCADE DYNAMICS OF A-T SUBJECTS 168

FIGURE 6.4: EARLY SACCADE VELOCITY OF A-T SUBJECTS 170

FIGURE 6.5: SIGNAL DEPENDENT NOISE 172

FIGURE 6.6: PREDICTIONS OF AN OPTIMAL FEEDBACK CONTROLLER. 175

Chapter 1: Introduction

Section 1.1: Redundancy in the selection of movement

Every day, we produce exquisitely precise movements of our eyes, body, and limbs. The brain quickly and seamlessly selects and executes these actions, allowing us to move smoothly in our environment and interact with the world. This apparent ease with which we make movements, however, belies the underlying complexity of the problem of motor control. At any given moment in time, the brain can select from a diversity of actions and produce a broad range of movements to change the state of the body. Yet, movements are generated rapidly, decisively, and accurately. In this work, we examined how the motor system selects among the breadth of possible behaviors to select and execute each of our actions.

Movements are presumably produced to place the body in a more desirable state or to achieve a desirable outcome. We think about these states and outcomes as having a value. For example, if I am thirsty, I might value states of my limb which bring my coffee mug to my mouth. In addition to the value of the outcome, the path I use to achieve that outcome also has an associated value. For example, the brain may desire to move the mug both quickly and accurately. A path which allows me to drink sooner is likely more desirable than one which takes longer. At the same time, fast movements that

may misplace the mug or result in spilling would be less desirable compared to slow movements that keep the liquid steady. The goals of moving to the mouth quickly and accurately are hence in conflict: to move the mug elsewhere may be faster; to reach quickly may be less accurate. In general, the selection of movement is a problem of optimization given conflicting values. Actions are presumably selected in a manner to balance these conflicting desires.

Internal values and desires are only a part of the puzzle, however. The selection of a desirable action is not limited to internally generated goals and constraints. External factors can also influence the selection of action. For instance, in the example above, I may be unaware of the amount of coffee in the mug, and so need to take that into account in deciding whether and how to move it – if the mug is full, more force will be required to lift it than if it is empty. The environment I am in can also affect the constraints on and value of my possible action – if flying in an airplane, turbulence might make it more likely to spill, increasing the value of slower, careful movements. Uncertainty, environmental constraints, and a changing world can further influence how movements are selected. To integrate these factors, the brain must predict and sense the state of the environment, and integrate this information in the selection of movements.

How can we disambiguate what and how factors are weighed in action selection? In practice, we cannot directly observe how goals and constraints are being evaluated by the nervous system. We can, however, observe the

consequences of the actions people make. By observing these consequences – how the eyes, body, and limbs move in time – we can draw inferences about the underlying goals. By applying external constraints or by exploiting interindividual differences in internal factors, we can identify the role of some of the factors in the selection of movement. In this work, we focus on the role of three of these factors: reward, error, and time.

Section 1.2: Tools to disambiguate goals in motor control

The study of motor control centers on the observation of movements produced by the brain and body. We studied two classes of movement: goal directed reaching movements and saccadic eye movements. In addition, we used several techniques to disambiguate the interactions between the goals underlying action selection. In Chapters 2, 3, and 4, we perturbed upper limb movements to probe the underlying control policy. In Chapter 5, we focused on differences between people in the way saccadic eye movements are generated. In Chapter 6, we studied people with a neurological disorder to understand how they, and healthy people in general, shifted their gaze.

Each of these techniques – the use of perturbations, the study of interindividual differences, and working with people with movement disorders – provided unique advantages towards understanding the way movements are selected and executed.

Perturbations

To study the policies underlying action selection, it is often helpful to perturb movements. In this work, we asked subjects to make goal directed movements towards a target. In some experiments, we altered the relationship between actions and the consequences of those actions by applying forces to the limb or manipulating the visual feedback participants received. Applying perturbations provided a critical tool for the study of movement in several ways.

First, it allowed us to observe what action is selected in the novel environment. This gave us the ability to test hypotheses about the factors contributing to the selection of action. For example, to test how the brain values reward, we could modulate when and how success was indicated for a movement. In a controlled experiment, we could observe if novel actions were selected in a way that took into account the changes in success feedback. To study the role of error, we could alter the relationship between actions and their sensory consequences, and observe how actions are selected when the distribution of errors is altered. Prior studies have used perturbations to provide a wealth of information about some of the factors involved in the selection of behavior. For example, they have demonstrated that the brain seeks to minimize prediction errors (Mazzoni and Krakauer, 2006) as well as produce accurate movements that result in success on the task (Izawa and Shadmehr, 2011).

Second, applying perturbations allowed us to observe the dynamics with which a new behavior is selected – the methods by which the brain achieves its preferred behavior. For example, prior work has demonstrated that when a perturbation is applied, the brain develops a representation, a model, of the environment (Thoroughman and Shadmehr, 2000; Wagner and Smith, 2008). This representation is used to select behaviors, and develops gradually as errors are experienced. In particular, the motor system learns from prediction error (Mazzoni and Krakauer, 2006; Taylor et al., 2014) – the difference between the expected and observed consequences of an action. The learning accumulates with time, and is represented in memory with multiple time scales (Smith et al., 2006; Kording et al., 2007). Memories of prior experiences persist (Smith et al., 2006; Kording et al., 2007; Pekny et al., 2011; Shmuelof et al., 2012), evidenced by signs of those experiences later in an experiment. The study of the dynamics of the response to a perturbation allows us to understand the time course and methods by which novel actions are selected when making goal-directed movements.

Interindividual differences

The study of interindividual differences in movement control also provides a powerful tool to understand the computations involved in the selection of behavior. Each of our brains likely adopts a similar general policy to choose an action, but dissimilarities in our bodies and personalities lead to subtle

distinctions in the way different people choose to act. Even healthy people vary in their impulsivity – the way they value rewards in time (Myerson and Green, 1995) – and in the variability of the errors their actions produce (Wu et al., 2014). We exploited these natural differences and examined how people who valued time more or who exhibited greater errors made movements.

The study of saccadic eye movements provides a unique arena to examine these differences, because people make these saccades in a characteristic way (Collewijn et al., 1988), termed the 'main sequence' of saccades (Bahill et al., 1975). The stereotypy of the main sequence has led to significant insight into some of the components that may drive the selection of the dynamics of saccades (Harris and Wolpert, 1998, 2006; Todorov and Jordan, 2002; Shadmehr et al., 2010). For example, these studies have shown that a hyperbolic cost of time provides a good fit to the dynamics observed in the main sequence (Shadmehr et al., 2010). Instead of studying the way an average individual moved, as in prior work, we examined the differences in the way individuals chose to make eye movements. We paired these measurements with estimation of the way each person valued reward in time or of the errors their actions produced. We then used mathematical formulations of the costs underlying decision making and movement selection to understand how each of these factors contributed to interindividual differences in motor control.

Movement disorders

Finally, working with people with neurological disorders provides a third powerful method towards understanding the factors involved in the control of movement. Studying people with these disorders provides an alternative mechanism to examine changes in the way movements are selected when the motor system is perturbed. For example, people with some neurological disorders exhibit systematic changes in the way they value reward (Frank et al., 2004) or time (Shadmehr et al., 2010). These changes are reflected in differences in the way behaviors are selected and movements are executed. Furthermore, in many disorders, the neural substrates of the disease are also known. Measurement of the movements produced by healthy people and people with movement disorders can provide unique insight into the complexities of movement control – how the brain selects actions in both health and disease.

Section 1.3: Scope of this work

In each chapter in this work, we made use of the diversity of actions available to solve a task to understand how actions are selected by the brain in movement control. We presented the motor system with a task with redundancy, where several actions can be used to be successful, and observed what solution people adopt, and how they achieved that solution. Our goal was to understand the costs and tradeoffs that are involved in how movements are

selected. In particular, we focused on the role that errors, rewards, and time play in action selection.

In Chapter 2, we eliminated error in "error-clamp" trials (Scheidt et al., 2000). These perturbations imposed additional redundancy in the task: subjects could produce many actions, all of which would result in the desired outcome, with no error. We found that eliminating error in this way resulted in a detection of a contextual change in a task and triggered a change in behavior. The contextual change was detected only when the distribution of errors in the task was altered. When a change was detected, we found that the short term history of training affected the ultimate, triggered behavior subjects selected in the error-clamped environment; the long term history of training affected how the ultimate behavior was achieved.

In Chapter 3, we also used error-clamp trials to impose task redundancy, but critically created a situation where subjects could produce many actions, all of which resulted in an undesired outcome. That is, instead of eliminating error, we presented subjects with persistent errors regardless of their action. We found that, despite their willingness to tolerate errors of comparable magnitude during training, subjects adopted an alternative mechanism of learning – exploratory behavior – in the error-clamp trials with persistent errors. Engagement of alternative mechanisms was prevented when, as in Chapter 2, we impaired the ability to detect a contextual change in the task. When change detection was

prevented, typical error-based learning mechanisms remained engaged in the clamp trials. Changes in the distribution of error, and not reward, were critical to triggering alternative mechanisms of learning.

In Chapter 4, we wanted to examine what behaviors subjects would produce when achieving task success (reward) was at odds with minimizing error in a task. Typically, these two goals can be achieved together: actions that minimize error also maximize reward. However, distinct neural bases are thought to be involved in learning from these two signals (Doya, 1999). We examined what behaviors were selected when these goals were placed in conflict. We found that prediction error provided a strong learning signal to the motor system that dominated behavior, even when it resulted in a reduction in reward. In the absence of prediction errors, the motor system did select behaviors to achieve reward. However, only when the prediction error signal was mild and the reward signal was strong did we observe a balance between the two systems of learning. Otherwise, sensory prediction error dominated the selection of actions.

In Chapter 5, we studied the eye movements of adults in a decision making task. We used a task where participants could be more successful – attain additional reward – if they were willing to wait longer in the task. That is, the task examined the tradeoff between time and reward in each participant. We wondered if the balance between these two costs in decision making would also

be reflected in the movements of participants. We found that people who were more willing to wait in the decision making task also produced slower movements, while those who were less willing to wait produced faster movements. That is, by examining the relationship between time and reward across individuals, we found evidence for a common cost of time in decision making and movement control.

Finally, in Chapter 6, we examined the eye movements of people with ataxia-telangiectasia (A-T), a neurodegenerative disorder, and of people without neurological problems. We found differences in the movements of people with A-T: when shifting their gaze, people with A-T made 2-6 movements, while control participants made only 1-2 movements. We wondered if this behavior could be explained by a tradeoff between a desire to be accurate (reduce endpoint error), but also achieve the target quickly (a cost of time). A-T subjects, on average, made movements with greater variability, and so making several smaller movements reduces endpoint error. We found that the behavior of both control and A-T subjects was well accounted for by this tradeoff – the selection of the number and gain of saccades reflected the noise properties of the oculomotor system, and a desire to minimize costs of both error and time.

Chapter 2: Decay of motor memories in the absence of error[1]

Section 2.1: Motivation

When motor commands produce unexpected results, the brain changes the commands on the subsequent trial (Thoroughman and Shadmehr, 2000). For example, when one reaches while holding an object that has novel dynamics (Shadmehr and Mussa-Ivaldi, 1994), the sensory consequences of motor commands are different than expected. This prediction error alters the motor commands on the subsequent trial, and the accumulation of this error-dependent learning, combined with repetition of the motor commands (Huang et al., 2011), produces a motor memory that can be recalled months later (Shadmehr and Brashers-Krug, 1997). The ability to protect and recall a memory can be strengthened by altering the perturbation schedule, amount of repetition, or type of feedback (Huang et al., 2011; Pekny et al., 2011; Shmuelof et al., 2012). Paradoxically, in one condition the motor memory appears transient: if training is followed by a block of error-clamp trials, trials in which errors are artificially eliminated, the motor output decays. That is, error produces a change in motor output, but absence of error also produces a change. The change in motor output in the absence of error has been taken as evidence that error-dependent

[1] The work in this chapter has been published in the Journal of Neuroscience: Vaswani PA, Shadmehr R (2013) Decay of motor memories in the absence of error. J Neurosci 33:7700–7709.

adaptation produces a motor memory that is inherently labile, exhibiting decay (Cheng and Sabes, 2006; Smith et al., 2006; Criscimagna-Hemminger and Shadmehr, 2008; Joiner and Smith, 2008; Zarahn et al., 2008; Galea et al., 2011). How is it that recently acquired motor memories exhibit decay in the absence of error over tens of trials, yet have long-term stability as exhibited by recall months later?

An important clue is an observation that was made by Scheidt et al. (2000). Subjects learned to compensate for a perturbation, and were then exposed to error-clamp trials. In most healthy subjects, in the absence of error the motor output decayed at the very onset of the error-clamp block, exhibiting zero lag (Fig. 5B in Scheidt et al. 2000). On rare occasions, however, motor output showed no decay for tens of trials: in one subject there was a 60 trial lag (Fig. 5C in Scheidt et al. 2000). Therefore, absence of error is not a sufficient condition for decay of motor output. Rather, the trial at which the decay begins may be probabilistic.

The aim of the work in this chapter was to better understand how actions are selected and why motor output decays in the absence of error. In our first experiment, we explored the question of the steady state (or endpoint) to which motor output decays. We found that this endpoint was never zero, but always a fraction of the motor memory that was last acquired. In our sample of subjects we also observed an occasional instance in which the decay started not at the

beginning of the error-clamp block, but after a lag. In our second and third experiments, we explored conditions that changed this lag. We found that the lag could be extended, sometimes indefinitely, by closely matching the statistics of movements in the error-clamp trials to the preceding training trials. These results suggest that motor memories do not passively decay in the absence of error, but are actively disengaged because the brain detects a change in the task.

Section 2.2: Methods

Fifty-five healthy right-handed subjects with no known neurological impairment participated in this study (age 23.1±4.6 years, 31 females). All subjects were naive to our apparatus, the paradigm, and the purpose of the experiment. The work in this chapter was approved by the Johns Hopkins University School of Medicine Institutional Review Board.

Subjects grasped the handle of a two-link robotic manipulandum with their right hand and made reaching or shooting movements, described below, for approximately 60 minutes. A screen covered the hand; veridical visual feedback (3 mm diameter cursor) was provided throughout the experiment. Subjects were permitted short breaks at defined points in the experiment. Position, velocity, and force at the handle were recorded at a rate of 200 Hz. During some trials, participants experienced a viscous curl field $\mathbf{f} = kB\dot{\mathbf{x}}$ (Figure 2.1A), where $\mathbf{f}$ is force on the hand, k is a constant describing the field strength, B is a viscosity

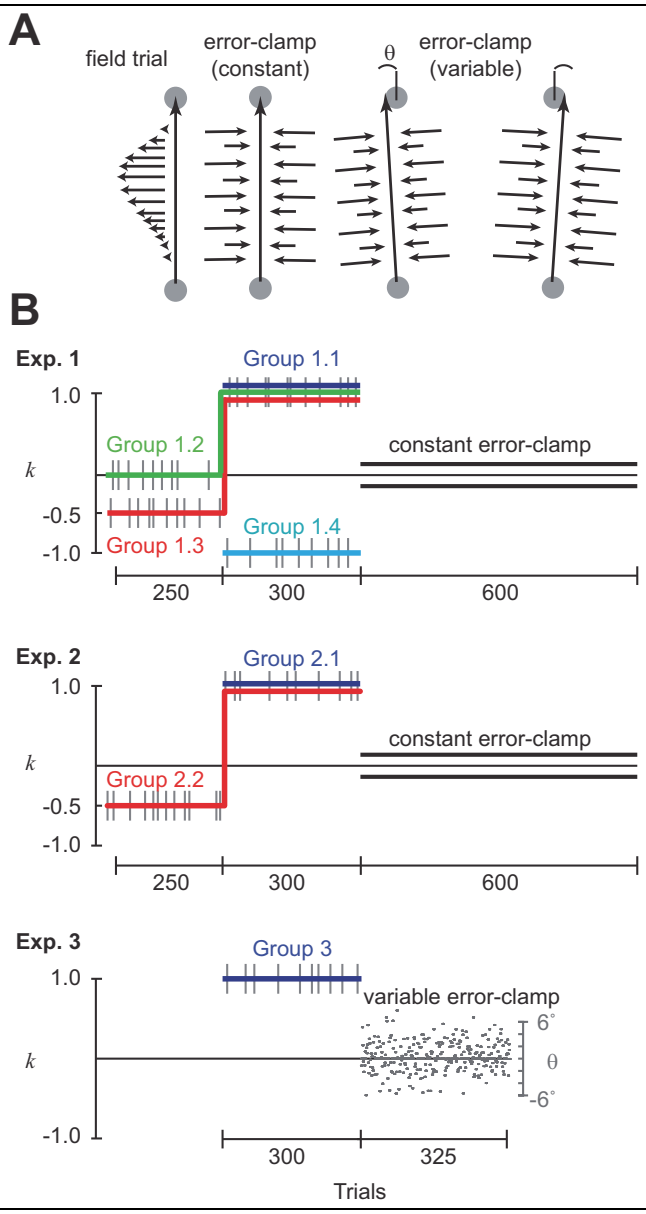

Figure 2.1: Experimental paradigm**.** **A**, Participants experienced three types of trials. In field trials, participants experienced a viscous curl field $\mathbf{f} = kB\dot{\mathbf{x}}$, where $\mathbf{f}$ is force on the hand, k is a constant describing the field strength, B is a viscosity matrix describing a counter-clockwise curl field, and $\dot{\mathbf{x}}$ is the hand velocity. In error-clamp trials, movements were constrained to a straight line by a stiff one dimensional spring. Constant error-clamp trials constrained movements in a line to the target. In variable error-clamp trials, the movement endpoint was a random variable θ that matched the natural variability in subjects' movement (Shmuelof et al., 2011). **B**, In Exp. 1, subjects made point to point reaching movements. In Exp. 2 and 3, subjects made shooting movements through and past the target. k indicates the field strength. Constant error clamp trials (thin grey lines) were interspersed randomly with 20% probability through the experiment. θ (grey points) indicates the movement endpoint in variable error clamp trials. Black parallel lines indicate constant error-clamp trials.

matrix describing a counter-clockwise curl field, $B = \begin{bmatrix} 0 & -15 \\ 15 & 0 \end{bmatrix}$ Ns/m, and $\dot{\mathbf{x}}$ is the hand velocity vector. In some trials, subjects experienced an error-clamp (Figure 2.1A) in which the movement was constrained to the line between the start position and an endpoint (usually the target, except in Experiment 3, see below) by a stiff, one dimensional spring (spring coefficient 6 kN/m, damping coefficient 250 Ns/m) (Scheidt et al., 2000). In these trials we were able to record the compensatory forces exerted by the subject onto the walls of the channel formed by the spring. We placed error-clamp trials with 20% probability randomly throughout each experiment.

Our aim was to understand why motor output decays in error-clamp trials. To answer this question, we began with a series of experiments in which we varied the history of the training (Exp. 1). We followed up with another series of experiments (Exps. 2 and 3) in which we tested the hypothesis that the decay occurs only if the brain detects a change from the field trials to error-clamp trials.

Experiment 1: Reaching with constant error-clamp

In Exp. 1, our aim was to determine the effect of the history of training on the rate, endpoint, and lag of the decay in motor output in the error-clamp trials. Participants (n = 24) made point to point reaching movements from a 1 cm diameter starting circle at approximately midline of the body to a 1 cm diameter

target 10 cm away. The start circle was visible throughout the experiment. Subjects were instructed to move 'rapidly and accurately' to the target once the target appeared and an auditory cue was played. Once the hand stopped at the target, feedback was provided. If the movement took too long (> 550 ms), the target turned blue. If the movement was too brief (< 450 ms), the target turned red. Subjects received a point and an 'explosion' of the target for movements between 450 and 550 ms. Auditory feedback was also provided: a 2000 Hz tone indicated success and a 200 Hz tone indicated failure. After each trial, the target was removed and the robot returned the subject's hand from the target to the start position.

Four groups of subjects participated in this experiment (Figure 2.1B). Group 1.1 (n = 6) trained only in field $k = 1$ (counter-clockwise field, 300 trials) without baseline training in null. Group 1.2 (n = 6) trained in null (k = 0, 250 trials) followed by 300 trials of $k = 1$ field training. Group 1.3 (n = 6) trained in $k = -0.5$ (250 trials) followed by 300 trials of training in $k = 1$. Group 1.4 (n = 6) trained only in field $k = -1$ without training in null. After completion of training in field, all subjects completed 600 error-clamp trials.

Experiments 2: Shooting with constant error-clamp

In order to test the hypothesis that change detection may be a critical component of decay of motor memories in error-clamp trials, we tried to make

the differences between movements in field trials and error-clamp trials less pronounced. In Experiment 2, subjects (n = 12) were asked to 'shoot through and past' the target into a virtual pillow positioned beyond the target. The cursor was turned off once the subject moved past the target. The goal duration was 150-250 ms. Feedback for movements which were too fast, too slow, or correct was identical to the reaching task in Exp. 1. If subjects struck the virtual pillow in the appropriate time but missed the target, the target turned pink, a 500 Hz tone was played, and the trial was considered unsuccessful. Group 2.1 (n = 6) trained only in field $k = 1$ (300 trials) without baseline training in null. Group 2.2 (n = 6) trained in $k = -0.5$ (250 trials) followed by 300 trials of training in $k = 1$. After completion of training in field, all subjects completed 600 error-clamp trials.

Experiments 3: Shooting with variable error-clamp

In Exp. 3, we attempted to make the transition from field to error-clamp block more difficult to detect by making the error-clamp trials have a variability that matched the variability of field trials. Subjects (n = 19) trained only in field $k = 1$ (300 trials) without baseline training in null. This was followed by a block of variable error-clamp trials (325 trials). In these trials, as in constant error clamp trials, the movement was constrained to a line by a stiff one dimensional spring between the start position and an endpoint (Figure 2.1A). The variable error-clamps were designed so that the endpoint of a given error-clamp trial was

a random variable θ with a distribution that matched the natural variability of the endpoint of subjects' movements (Figure 2.1C) (Shmuelof et al., 2011). The distribution of this random variable was selected based on the average mean and variance of the distribution of the movement endpoint of subjects in Experiment 2 at the end of training. The angular deviation θ was drawn from a normal distribution with the following mean and variance: $\theta \sim N(-0.2°, 2.6°)$. Note that the small non-zero bias in the mean angular deviation of these error-clamp trials is equivalent to a 0.03 cm lateral deviation to the right at the end of the movement, well within the target width of 1 cm (5.7°). As in Experiments 1 and 2, error-clamp trials directly to the target were interspersed with 20% probability throughout the experiment.

Adaptation Index

We quantified performance via an adaptation index during error-clamp trials (Smith et al., 2006). The lateral force produced during an error-clamp trial was regressed onto the ideal compensatory force profile for a counter clockwise field with $k = 1$. The adaptation index was zero if the forces were uncorrelated and one if they were perfectly correlated. For example, perfect compensation for a counter clockwise field ($k = 1$) results in an adaptation index of 1; perfect compensation for the weak clockwise field ($k = -0.5$) results in an adaptation index of -0.5.

Decay of force in error-clamp trials

In all experiments, we assessed whether subjects reduced their motor output immediately upon entering the block of error-clamp trials or with a lag. The adaptation index of subjects in the last 50 trials of training and during the final error-clamp block were fit using a nonlinear least squares fit to a lagged exponential decay $f(\tau)$.

$$f(\tau) = \begin{cases} a & \tau \le \lambda \\ (a-b) * \exp\left(-\beta(\tau-\lambda)\right) + b & \tau > \lambda \end{cases} \qquad \text{Eq. 2.1}$$

In Eq. 2.1, τ is the trial number in the error-clamp block, a is adaptation index the subject achieved in field trials, b is the endpoint of the decay, β is the rate of decay, and λ is the lag.

$$\begin{aligned} a &\in [0.5, 1.5] \\ b &\in [-0.3, 0.3] \\ \beta &\in [0.005, 0.5] \\ \lambda &\ge 0 \end{aligned}$$

Statistics

All statistical analyses in this chapter were conducted using Matlab (R2012a, The Mathworks) or SPSS (V21, IBM). Unless otherwise indicated, we used Student's t-test, ANOVA, and repeated measures ANOVA with Greenhouse-Geisser correction when Mauchly's test of sphericity failed. One-

tailed tests were used where indicated when there was an *a priori* expectation of a directional effect.

Section 2.3: Results

Effect of history of training on rate and endpoint of decay

We began by asking whether the decay of motor output that typically takes place in error-clamp trials is influenced by initial training in a null field. In all previous publications of which we are aware, volunteers were initially trained in a null or baseline condition in which reaches were performed without a perturbation. We wondered whether this pre-training affected the motor output in the error-clamp block. If so, training subjects with a different baseline, or without baseline training, might change the endpoint of the decay.

To answer this question, we compared performance of two groups of subjects. Group 1.1 trained for 300 trials in a counterclockwise field (Figure 2.1B) but received no prior training in the null field (i.e., field was at full strength from the very first trial). In contrast, Group 1.2 trained for 250 trials in a null field, and was then exposed to the counterclockwise field (Figure 2.1B). Both groups experienced a block of error-clamp trials following field training. Our measure of motor output was an adaptation index, which quantified force produced in error-clamp trials as a function of ideal forces needed to compensate for a counterclockwise force field. To assess performance we considered the mean

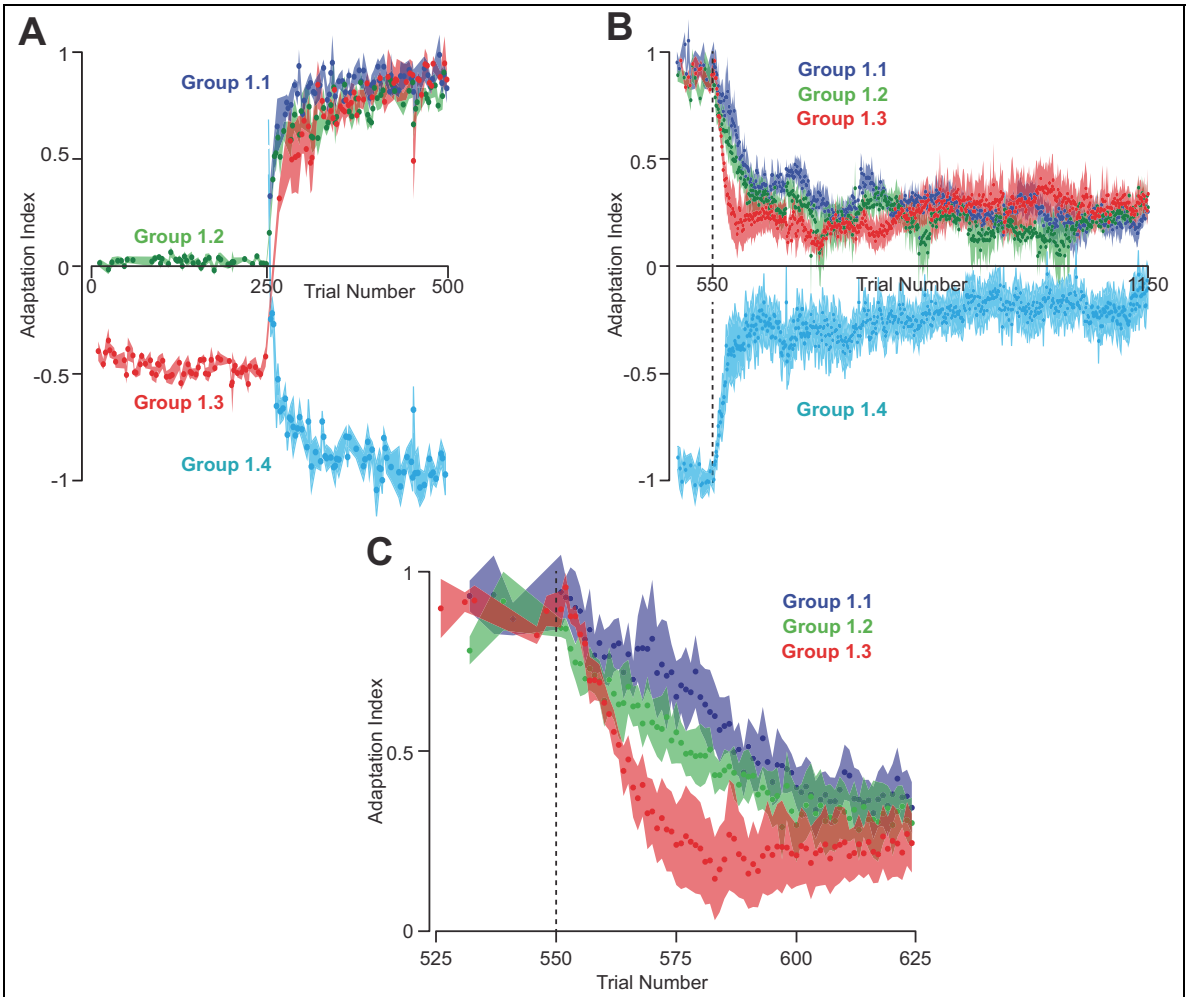

Figure 2.2: Learning and decay in reaching movements in Experiment 1. Motor output was quantified via an adaptation index. The adaptation index represents the linear regression of the ideal compensatory force profile onto the actual force profile measured in error-clamp trials. A, Time course of the change in the adaptation index during baseline training and adaptation and B, in final error-clamp trials. C, Initial decay. Data are mean±SEM across all subjects.

adaptation index at the end of field training (trials 500 to 550) and in the final 50 error-clamp trials in the experiment (trials 1100 to 1150). We found that by end of field training, performance in the two groups was comparable (Figure 2.2A and Figure 2.2B, adaptation index, mean±SEM, Group 1.1: 0.93±0.04, Group 1.2: 0.86±0.03, two-tailed t-test $p = 0.17$). The motor outputs decayed rapidly over the course of about 75 error-clamp trials (Figure 2.2C), but showed no further decay

over the next 500 trials (Figure 2.2B). The decay endpoints were similar in the two groups (Group 1.1: 0.21±0.05, Group 1.2: 0.26±0.03, $p = 0.41$). Furthermore, after 600 error-clamp trials, the motor outputs in both groups were still significantly greater than zero (one-tailed t-test, Group 1.1: $p = 0.005$, Group 1.2: $p = 2E\text{-}4$). Therefore, motor output decayed in the error-clamp trials whether or not subjects were exposed to the null field. The decay endpoint was not zero, but a fraction of the motor output attained during training.

To further investigate the factors that contributed to the decay endpoint, we recruited a new group. Group 1.3 first trained in a clockwise field, and then in a counterclockwise field (Figure 2.1B). Subjects learned to compensate for the counterclockwise field as well as subjects in Groups 1.1 and 1.2 (adaptation index, mean±SEM, Group 1.3: 0.89±0.01; ANOVA, effect of group on training: $p = 0.25$). However, in error-clamp trials, the motor output of Group 1.3 declined somewhat faster than Groups 1.1 and 1.2 (repeated measures ANOVA for trials 550 to 625, significant within-subject interaction effect between group and trial number: $p = 0.02$), but had an endpoint that was similar to the two other groups (adaptation index, mean±SEM, Group 1.3: 0.28±0.07; ANOVA, effect of group on endpoint: $p = 0.57$). This suggested that no matter which sequence of fields subjects trained in, decay endpoint was always a fraction of the last field to which they had been exposed.

To check the validity of this conclusion, we a recruited a new group and tested them in a field opposite to all previous groups. Group 1.4 never trained in a null field and was only exposed to a clockwise field (Figure 2.1B). In error-clamp trials, the motor output of Group 1.4 also declined (Figure 2.2B). Like other groups, the decay endpoint for Group 1.4 was a fraction of the motor output learned in the last set of field trials (adaptation index, mean±SEM, Group 1.4: -0.19±0.09, one-tailed t-test $p = 0.04$). We conducted an ANOVA to determine the effect of the baseline field (clockwise, null, or none) and training field (clockwise or counterclockwise) on the decay endpoint in these four groups. There was a significant effect of the training field ($p = 3E\text{-}4$) but no effect of baseline field ($p = 0.67$).

If subjects are reducing their motor output in error-clamp trials in order to minimize the effort of their movements (Emken et al., 2007), or because a motor memory is passively being forgotten from trial to trial (Smith et al., 2006), we would expect decay endpoints to be zero. In contrast, our results show that decay endpoint is a non-zero fraction of the last motor output learned during training. Furthermore, our results show that whereas the rate of decay depends on the sequence of fields that have been learned, the decay endpoint depends only on the final field that was learned.

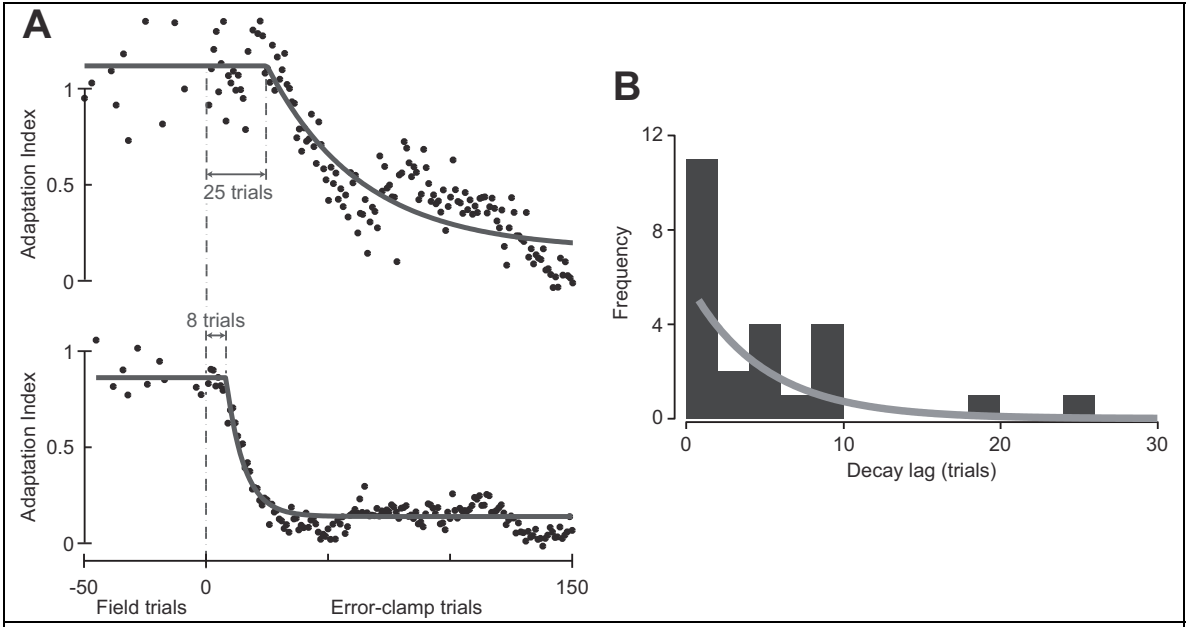

Figure 2.3: Lag to decay in reaching movements in Experiment 1. A, Time course of the change in adaptation index in error-clamp trials for two example subjects (top, Group 1.7; bottom, Group 1.1). A delayed exponential was fit to the data. B, Distribution of lags to the start of the decay for all subjects.

A lag to the start of the decay

In closely inspecting our data we noticed that some subjects did not show a decay of their motor output at the onset of the error-clamp trials. Rather, there were some subjects who displayed a lag to the start of the decay, as shown by the data from two subjects in Figure 2.3A. To quantify this pattern, for each subject we fit a delayed exponential (Eq. 2.1) to the final 50 trials of adaptation and the 600 trials of the error-clamp trials and estimated their lag λ (in units of trials). Across subjects, λ is a random variable with the distribution shown in Figure 2.3B. We found that on average there was a small, but significant lag of 4.8±1.3 trials (mean±SEM, one-tailed t-test p = 7E-4). The lag was not different between groups (ANOVA, effect of group on delay, p = 0.98).

We next wanted to assess the distribution of the lags in this experiment. If the initiation of decay is a probabilistic event occurring independently on a trial-to-trial basis, we would expect the distribution of lags to be exponential. We computed the maximum likelihood fit of the decay lag λ to normal, Poisson, gamma, and exponential probability distributions, and used the Akaike Information Criterion (AIC) to assess the relative quality of the fits. In order to fit a Poisson distribution, the lags were rounded to the nearest integer. In order to fit a gamma distribution, which has a probability of zero at zero, the distribution of the first trial of decay (i.e. lag plus one) was used. We found that the decay lag λ was best fit by an exponential probability distribution $P(\lambda) = \frac{1}{\beta}\exp\left(-\frac{\lambda}{\beta}\right)$, resulting in $\beta = 4.8$ trials (AIC, exponential:125, normal:160, Poisson:238, gamma: 136). The 95% confidence interval of β was [3.3, 7.5] trials (goodness of fit log-likelihood of -62), further confirming that the mean of the distribution of the lags is quite likely to be greater than zero.

The data from Exp. 1 (Figure 2.3B) hinted that the start of the decay may be a probabilistic event: for many subjects the decay started at the first error-clamp trial, but for others started a few trials later. To explain this probabilistic behavior, we hypothesized that the decay of motor output may have been initiated when the brain detected a change from the field block to the error-clamp block of trials.

To better understand what may be changing between these two block of trials, we quantified performance of each subject at the end of the force field block of trials (FF, last 20 trials) and at the beginning of the error-clamp block of trials (EC, first 20 trials). The data are summarized for each subject in the left column of Figure 2.4. We found that when subjects transitioned from field to error-clamp trials in Exp. 1, the error-clamp trials made the movements artificially more straight (deviation of trajectories from a straight line, mean±SEM, FF: 18±2 cm^2, EC:0.15±0.02 cm^2, one-tailed t-test of FF vs. EC, p = 3E-7, Figure 2.4A), more similar from trial to trial (inter-trial correlation of velocity time-series, FF:0.41±0.03, EC:0.56±0.03, p = 2E-4, Figure 2.4B), more successful in terms of acquiring reward (probability of reward, FF:0.43±0.04, EC:0.81±0.02, p = 1E-9, Figure 2.4C), and shorter in duration due to a need to make fewer corrections (movement duration, FF:577±13 ms, EC:479±12 ms, p = 1E-7, Figure 2.4D). Therefore, at the onset of the error-clamp block, there was a significant change in movement kinematics. It seems reasonable that the motor system detected some of these changes. We hypothesized that if detection of these changes is partly responsible for the non-zero lag that we observed in the initiation of decay, then reducing the differences between field and error-clamp trials should increase the lag. We pursued this idea in Experiments 2 and 3.

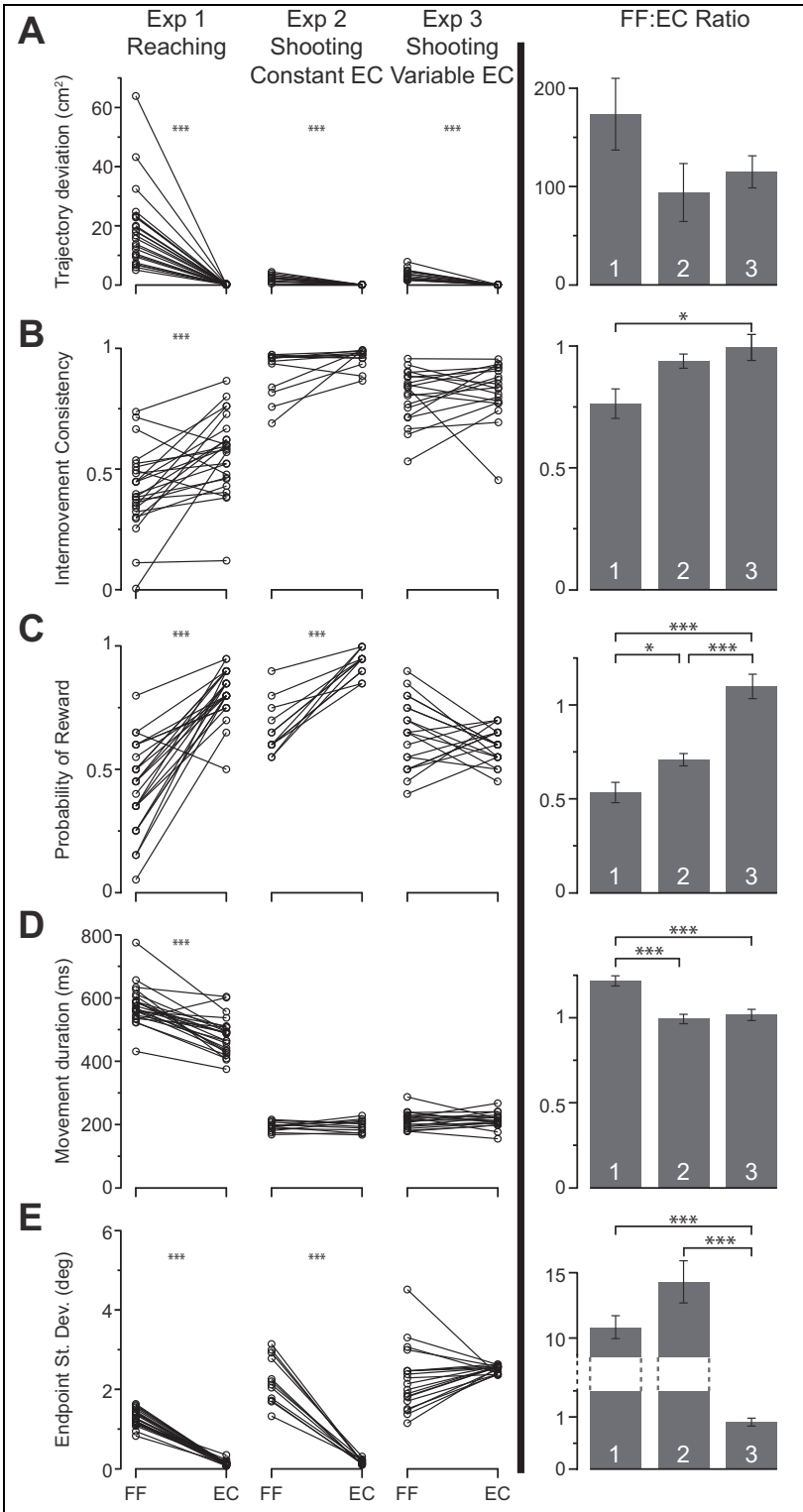

Figure 2.4: Change in behavior between the end of the force field (FF, last 20 trials) block of trials and the beginning of the error-clamp (EC, last 20 trials) block of trials. A, Trajectory deviation: sum of the squared deviation of each point in the trajectory from the straight line connecting the start to end positions. B, Inter-movement consistency: correlation between the velocity time-series from one trial to the next (Shadmehr and Mussa-Ivaldi, 1994). C, Probability of reward. D, Movement duration. E, Endpoint angle: the angle between the line from start to the target and the line connecting the start and end positions. *Left column*: the change in task parameters for reaching movements in Exp. 1, shooting movements with constant error-clamp in Exp. 2, and shooting movements with variable error-clamp in Exp. 3. Lines are the mean (A-D) or standard deviation (E) of the measure for each subject. One tailed t-test that consistency and reward increase, and deviation, duration, and endpoint variability decrease in EC trials. *Right column*: The ratio of the value of each measure in FF to its value in EC trials (FF:EC ratio). A ratio of 1 indicates no change. Data are mean±SEM across all subjects. One tailed t-test that FF:EC ratios of consistency and reward increase, and deviation, duration, and endpoint variability decrease, in later experiments. *p < 0.05 ***p < 0.001.

Increasing the lag

In Exp. 2 (Figure 2.1B), we asked subjects to make a shooting movement so that the hand went through and past the target and then hit a virtual pillow. In this paradigm they did not have to correct their movements or stop at the target. Our logic was that because a shooting movement to a target at a given distance is shorter in duration and is straighter than a reaching movement to the same target, in the shooting task performance at the end of the adaptation block might be more similar to the error-clamp block, making it harder for the brain to detect a change. Indeed, we found that performance measures in the two blocks of trials were now more similar. The data for the shooting task are summarized for the last 20 field trials and the first 20 error-clamp trials in column labeled Exp. 2 in Figure 2.4. To quantify the change in performance from the end of field trials to start of error-clamp trials, we computed the ratio of the value of each measure in the field trials to its value in error-clamp trials (FF:EC ratio, right column of Figure 2.4). We found that the change in performance from field trials to error-clamp trials was less pronounced in most measures in the shooting task as compared to the reaching task (one tailed t-test, straightness $p = 0.08$; consistency $p = 0.03$; reward $p = 0.02$; movement time $p = 1E\text{-}5$). Therefore, in the shooting paradigm (Exp. 2), movements in the field and error-clamp blocks of trials were more similar than the reaching movements in the same two blocks of trials (Exp. 1). Did this increased similarity affect the decay lags?

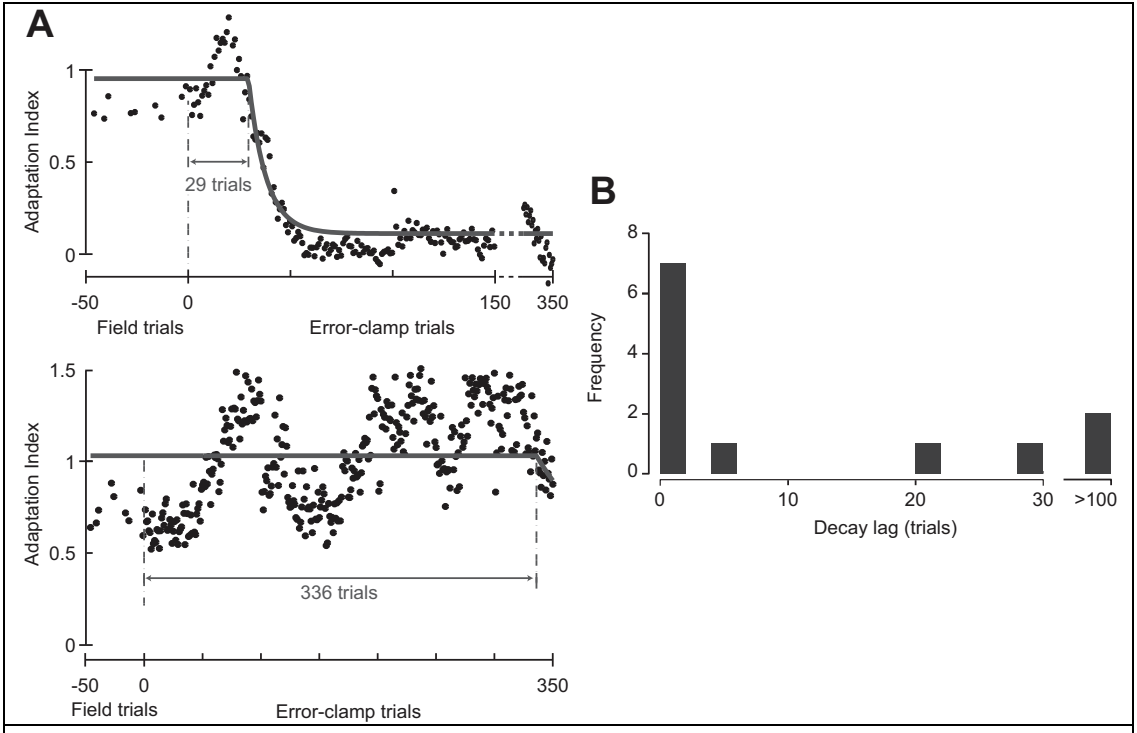

Figure 2.5: Lag to decay in shooting movements with constant error-clamp trials in Experiment 2. A, Time course of the change in adaptation index in error-clamp trials for two example subjects. Both subjects were in Group 2.2. A delayed exponential was fit to the data. B, Distribution of lags to the start of the decay for all subjects.

We quantified the lag to decay of motor output in the shooting experiment (Exp. 2). Data from two subjects are shown in Figure 2.5A. In one of these subjects we see a lag of 25 trials, whereas in the other subject the motor output shows no decay for over 300 trials. Interestingly, most subjects, including those who show decay in motor output, did not report observing any changes in the dynamics of the manipulandum when asked after the experiment was complete. However, when asked about her perceptions of the experiment, the latter subject (Figure 2.5A, bottom) reported noticing that the manipulandum was pushing her to the left, and then began to make her movements straight; however she showed no evidence of decay. Together, these suggest that explicit awareness of the

change in the task may be independent of the motor system's detection of a change.

The distribution of lags for the entire population of subjects in Exp. 2 is shown in Figure 2.5B. We fit the data in Figure 2.5B to an exponential probability distribution $P(\lambda) = \frac{1}{\beta}\exp\left(-\frac{\lambda}{\beta}\right)$, and found that $\beta = 44$ trials (95% confidence interval of [27, 86] trials, goodness of fit log-likelihood of -57). In comparing the lag distributions in Exp. 1 and Exp. 2, we find that the mean lag is significantly greater in Exp. 2 than in Exp. 1 (mean±SEM, Exp. 2: 44.3±28.8 trials, one-trailed t-test p = 0.03). Furthermore, we find that the mean of the lag distribution β has increased by an order of magnitude, and the 95% confidence intervals are well separated. In summary, as we increased the similarity between the movements in the field and error-clamp blocks of trials, we found an increase in the decay lags.

To pursue this idea further, we considered an experiment (Exp. 3) in which we attempted to delay the decay indefinitely by making the characteristics of movements in the error-clamp more closely match the statistics of movements that subjects produced at the end of training in the field trials. To do so, we focused on two variables: probability of success, a variable that was previously shown by Pekny et al. (2011) to act as a cue for the brain to detect a change in the properties of the task; and the trial-to-trial variance of movements. In the

column marked Exp. 2 in Figure 2.4, we see that probability of reward increased significantly from the end of field trials to start of the error-clamp block (one-tailed t-test, p = 2E-6). Similarly, the standard deviation of movement endpoints decreased significantly during these two periods (one-tailed t-test, p = 5E-8). In Exp. 3, we attempted to make these two measures more similar in the field and error-clamp trials. To achieve this goal, we introduced trial to trial variance in the error-clamp trials that matched the variance of movement endpoints at the end of training in the field trials.

In the final field trials of Exp. 2, subjects on average had a movement endpoint 0.2° to the right of center of the target (target was 5.7° in width), with an average within-subject standard deviation of 2.6°. In Exp. 3, we again considered shooting movements, but we varied the endpoint of the line to which subjects were constrained in the error-clamp trials to match the variability of movement endpoints of subjects in the final field trials of Exp. 2. The resulting movements are summarized in the column marked Exp. 3 in Figure 2.4. We found that endpoint standard deviation, movement duration, probability of reward, and inter-movement consistency were now similar at the end of the field block and the beginning of the error-clamp block (FF:EC ratio significantly different than 1, two-tailed t-test, p = 0.21, p = 0.65, p = 0.15, and p = 0.93 respectively). Therefore, by making the error-clamp trials slightly variable in

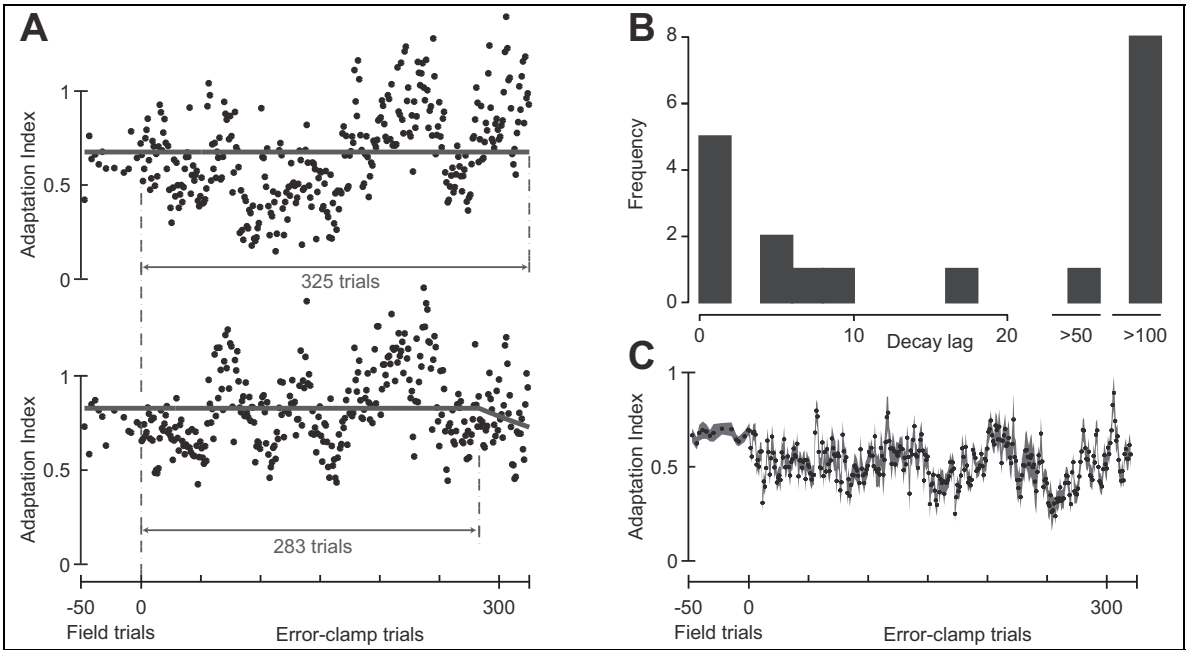

Figure 2.6: Lag to decay with variable error-clamp trials in Experiment 3. A, Time course of the change in adaptation index in error-clamp trials for two example subjects. Both subjects were in Group 3.1. A delayed exponential was fit to the data. B, Distribution of lags to the start of the decay for all subjects. C, Time course of the change in adaptation index in error-clamp trials for all subjects. Data are mean±SEM across subjects.

Exp. 3, we were better able to match the statistics of field and error-clamp block of trials.

Interestingly, a large fraction of subjects in Exp. 3 showed no appreciable decay in their motor output, even after 100 or more error-clamp trials. Data from two of these subjects are shown in Figure 2.6A. As a population, we observed a mean decay lag of 96.1±27 trials (mean±SEM) (Figure 2.6B). Eight of the 19 subjects in this group had a lag of greater than 100 trials. Indeed, as a group there was no evidence of decay in the error-clamp trials (Figure 2.6C). The lackof decay cannot be due to the small bias in the endpoint distribution of the movements in the error-clamp block: the bias in the mean angular deviation of

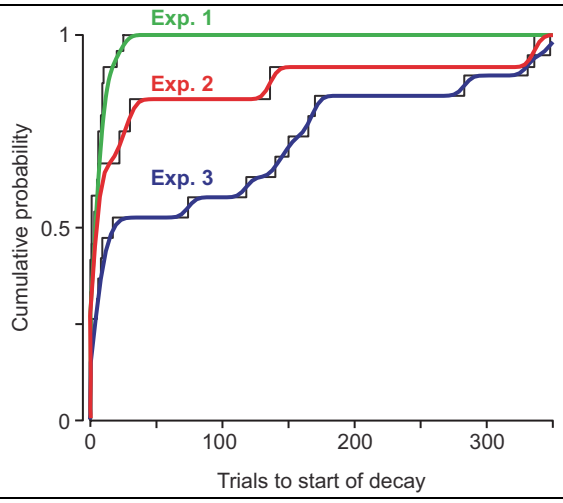

Figure 2.7: Cumulative probability distribution of the lag to decay in Experiments 1, 2, and 3. An estimated cumulative probability density estimate for each experiment (in color) is overlaid on the data (black).

these error-clamp trials is equivalent to a 0.3 mm lateral deviation to the right at the end of the movement, well within the target width of 1 cm. Compensation of this bias by the subject would require 0.21 N of force to the left at the end of the movement. Subjects instead maintain a force of 6.7±1.0 N to the right

One way to compare the decay lags in the various experiments is to consider the cumulative probability distribution of lags in each experiment (Figure 2.7). These distributions allow us to estimate at what trial we can say with some reasonable certainty that the decay will have started in most subjects. As we moved from reaching to shooting, the decay lag (with 75% certainty) in the error-clamp block shifted from 6 trials in Exp. 1 to 22 trials in Exp. 2. When we made the error-clamp trials more closely resemble field trials, the decay lag shifted (with 75% certainty) from 22 trials in Exp. 2 to 165 trials in Exp. 3. Of course, at 95% certainty these numbers are even farther apart. Therefore, the

start of decay in motor output is a probabilistic event that can be manipulated by altering the properties of error-clamp trials.

Predicting an individual's lag

We next considered whether the changes in movement kinematics from the field to error-clamp block of trials can be used to predict the lag to decay for individual subjects. For each experiment, we used a multiple linear regression model, with the FF:EC ratio of each subject's straightness, consistency, reward, movement time, and endpoint variability as predictors of their lag. We also included the mean and variability of the adaptation index at the end of field training as additional predictors, and a constant term. For each experiment, the value of these parameters did not linearly predict the lag to decay (Experiment 1: $R^2 = 0.17$, F-test $p = 0.854$; Experiment 2: $R^2 = 0.62$, F-test $p = 0.556$; Experiment 3: $R^2 = 0.22$, F-test $p = 0.856$). This suggests that the motor system of different subjects may be weighing these parameters differently, so a single linear model for all subjects is not appropriate. Alternatively, other, related parameters may be driving the process of change detection, or these parameters may have a nonlinear effect.

Section 2.4: Discussion

A few hundred movements with a novel tool can produce a motor memory that can persist for days, weeks, or months after training (Shadmehr

and Brashers-Krug, 1997; Joiner and Smith, 2008). Yet, motor memories have been considered inherently transient – that is, in the absence of an error signal, the motor output that was acquired during training decays immediately and automatically. This decay has been attributed to a number of possible processes, including trial to trial forgetting (Smith et al., 2006; Criscimagna-Hemminger and Shadmehr, 2008; Joiner and Smith, 2008; Galea et al., 2011) and optimization of effort (Emken et al., 2007; Ganesh et al., 2010). Our results challenge these views and suggest that decay is not forgetting or optimization in the absence of error, but a reflection of de-instantiation of a component of motor memory.

We considered a standard reach task in a force field and made three observations: First, decay of motor output in error-clamp block of trials was not to zero, but to an endpoint that was a non-zero fraction of the motor output in the last field subjects trained in. This observation was independent of the long term history of training. Second, the rate of decay was biased by the history of training, as prior training in null or the opposite field accelerated the rates of decay. Third, there was a lag to the initiation of decay. The lag was a probabilistic variable that differed among subjects: occasionally a subject would show little or no decay for many error-clamp trials, and then initiate decay. We were able to extend this lag significantly by making error-clamp trials more closely resemble field trials during learning.

Our results challenge three assumptions in current models of motor learning. The first assumption is that decay or removal of a memory in absence of error is an inherent part of motor memories, often attributed to a process of forgetting (Cheng and Sabes, 2006; Smith et al., 2006; Kording et al., 2007; Zarahn et al., 2008). While others have shown that reinforced memories persist and can be recalled later (Pekny et al., 2011; Shmuelof et al., 2012), our results suggest that the reduction in motor output in error-clamp trials is tied to a probabilistic event that depends on change-detection in the parameters of the task. Once a change is detected, decay follows, suggesting that decay is a form of de-instantiation and not an inherent part of the acquired memory, i.e., an active process that is initiated after an event is detected. We speculate that when the motor system detects a change in the parameters of the task, it de-instantiates part of the motor memory, which behaviorally masquerades as decay.

Second, our results are inconsistent with the assumption that decay is due to minimization of effort, a hypothesis proposed by Emken et al. (2007). In this case, one would expect the endpoint of decay to be independent of the training schedule. Here, we found that the endpoint of decay was a non-zero fraction of the last field subjects learned. For example, subjects who most recently trained in a counterclockwise field maintained a significant component of their recently learned motor output for 600 trials, even though the training was only for 300

trials, while subjects who most recently trained in a clockwise field maintained a significant component in the opposite direction.

Third, our results suggest that retention of a motor memory, as assayed in movements in the absence of error, consists of not one but multiple processes: a process that detects a change in context, behaviorally assayed via the decay lag; a process that de-instantiates a component of the memory, behaviorally assayed via the rate of decay; and a component of the memory that appears immune to this process of de-instantiation, behaviorally assayed via the endpoint of decay.

We interpret our results as follows. Motor memories consist of multiple components: one component that can be easily de-instantiated as task parameters change, and another component that is resistant to de-instantiation, expressed as decay endpoint. A recent brain stimulation study suggests a role for the primary motor cortex (M1) in the persistent component of motor memory, i.e., the endpoint of decay. Galea et al. (2011) trained subjects in a visuomotor rotation task. After subjects learned to compensate for the visual perturbation, they made movements in a series of trials where no visual feedback was presented – akin to error-clamp trials in force field learning. Galea et al. (2011) applied transcranial direct current stimulation (tDCS) to M1 and found that the stimulation had no effect on rates of acquisition as compared to a sham group. However, M1 tDCS significantly elevated the endpoint of the decay. This suggests a role for M1 in the persistent component of memories, assayed by the endpoint, at least for a

visuomotor rotation task. As M1 has been shown to be involved in use-dependent plasticity (Bütefisch et al., 2004), it is plausible this memory may develop in M1 through repetition and reinforcement.

Recently, Huang et al. (2011) suggested that learning to compensate for a perturbation installs a motor memory that may have three components: a forward model that associates motor commands with their sensory consequences and learns from prediction errors (Izawa et al., 2012; Schaefer et al., 2012), a bias in motor output that develops because of repetition of the motor commands (Diedrichsen et al., 2010; Verstynen and Sabes, 2011), and a reinforcement-dependent bias that develops because of reward dependent association between stimuli and the resulting motor commands (Huang et al., 2011; Izawa et al., 2012; Shmuelof et al., 2012). One may speculate that the decay endpoint in our data is the motor output that is being expressed by one of these processes, perhaps the process that learns from repetition and/or reinforcement.

We observed that the rate of decay was modulated by the history of training (Exp. 1). Pekny et al (2011) demonstrated that in error-clamp trials, when a change was inserted by manipulating reward feedback, prior motor memories were transiently recalled. The magnitude of recall was comparable to the magnitude of the persistent component of memory in that work. We speculate that the persistent component of memories is briefly recalled when a change is detected, resulting in the changes in rates of decay. Alternatively,

Huang and Shadmehr (2009) showed that the statistics of the history of training played a role in the subsequent rate of decay in error-clamp trials. While this finding cannot explain the recall of memory in Pekny et al. (2011), in our work it may be that Groups 1.2 and 1.3 have experienced larger errors than Group 1.1, and so decay faster. Once a change is detected, a combination of recall of previous memories and sensitivity to task statistics likely play a role in the process of de-instantiation.

We hypothesized that the similarity between field and error-clamp trials was a key factor in detecting a change, and indeed showed that by making these trials more similar, the lag to decay could be extended. However, one limitation of our work is that, using a linear model, we could not explain why a given subject produced their particular lag in a given experiment. That is, while our interpretations are likely true in a probabilistic sense, we have not shown that subjects who show greater lag in a given experiment are also those whose movements are more similar in the field and error-clamp blocks. It is possible that individual subjects weigh different kinematic and task factors differently, and so a single model across subjects is not appropriate; or these or other parameters may have a nonlinear effect.

We also observed that, in Exp. 2 and Exp. 3, subjects had trial-to-trial variability in their behavior during the error clamp block prior to the onset of decay. In Exp. 3 this may be, in part, a result of learning from the errors

presented in the variable error-clamp trials. However, learning from error cannot explain the variability seen in the constant error-clamp trials in Exp. 2, as no errors are presented. The low frequency of that variability in Exp. 2 suggests the behavior is not simply noise in the motor output. Subjects may be exploring their environment, or the variability may be a manifestation of the instability of the component of the motor memory that will be de-instantiated.

In summary, we offer evidence that adaptation installs two distinct types of motor memory that guide the selection of actions: one that is actively deinstantiated upon detection of change in the environment, and one that persists despite that detection.

Chapter 3: Exploratory escape from persistent errors[2]

Section 3.1: Motivation

When a motor command is produced but the result is other than expected, the brain partially compensates for the error by altering the commands on the next attempt (Shadmehr and Mussa-Ivaldi, 1994; Krakauer et al., 2000; Thoroughman and Shadmehr, 2000). As a result, when a perturbation is presented repeatedly, the changes in motor commands accumulate, largely compensating for the perturbation. A puzzling feature of this process of adaptation is that the compensation is often incomplete: after many trials of training, subjects still exhibit small, sustained errors in their performance (Kagerer et al., 1997; Krakauer et al., 2000, 2005, 2006; Miall et al., 2004; Tseng et al., 2007; Rabe et al., 2009; Shabbott and Sainburg, 2010; Galea et al., 2011; Taylor et al., 2014). It appears that even with extended training there are persistent steady-state errors that the brain does not correct for. Why should this be?

State-space models of learning provide a mathematical description of adaptation that can account for these persistent non-zero residual errors (Cheng and Sabes, 2006; Smith et al., 2006; Kording et al., 2007). In these models, it is

[2] The work in this chapter is in submission: Vaswani PA, Shmuelof L, Haith AM, Delnicki RJ, Huang VS, Mazzoni P, Shadmehr R, Krakauer JW. Exploratory escape from persistent residual errors in motor adaptation tasks.

assumed that the brain learns to estimate the state of the environment, updating its estimate after each trial based on the experienced error. In addition, the estimated state partially reverts toward baseline after each trial. That is, error-dependent learning is balanced with trial-to-trial forgetting (or reversion toward a baseline state). These two opposing effects eventually reach equilibrium in which learning from error balances reversion toward baseline. This trial-to-trial reversion was thought to be directly observable by using trials in which errors are constrained to zero, called error-clamp trials (Scheidt et al., 2000; Criscimagna-Hemminger and Shadmehr, 2008; Pekny et al., 2011; Kitago et al., 2013; Ingram et al., 2013).

Recent work, however, has revealed that reversion to baseline in adaptation paradigms is not inevitable. Close inspection of behavior in error-clamp trials reveals that subjects persist in their asymptotic actions over several trials before beginning to revert to baseline (see Chapter 2, Vaswani and Shadmehr, 2013). Reversion to baseline in error-clamp trials can be prevented altogether by reinforcing an action performed at asymptote, or by imposing variable errors in clamp trials (Shmuelof et al., 2012; Vaswani and Shadmehr, 2013). If reversion to baseline is not obligatory, why is it that subjects cannot close their residual error?

Shmuelof et al. (2012) previously suggested that the attenuation of reversion to baseline reflected the engagement of an alternative reinforcement-

based learning process, which is usually mostly suppressed in regular adaptation paradigms. This alternative process deviates from the dynamics of the state-space model and, in particular, may not be susceptible to reversion to baseline. Here, we asked whether such an alternative learning process could also be used to overcome the persistent residual errors seen during adaptation. We used a novel kind of error-clamp that imposed constant, small *non-zero* errors. We posited that the ensuing de-correlation of visual feedback from a subjects' actions, might create sufficient of a contextual change (Vaswani and Shadmehr, 2013) to provoke subjects to break free from error-based learning and attempt to close the residual error with another kind of learning.

Section 3.2: Methods

Participants

Fifty seven healthy, right handed subjects (34 females, 20-41 years of age) participated in the experiments. All subjects were naïve to the purpose of the experiment and were paid to participate. Experiments in this chapter took place at Columbia University and were approved by the Columbia University Institutional Review Board.

Paradigm

Subjects sat at a table with their right hand supported on a lightweight sled. Air jets in the sled generated air cushions that facilitated frictionless planar

movements. Subjects could not see their hand, but were provided with continuous veridical visual feedback throughout the experiment. Custom routines controlled the real-time visual display.

Subjects performed movements with their right arm from a starting circle to a circular target (0.5 cm diameter) positioned 8 cm away at 135° (Figure 3.1A). Hand and arm position were recorded at 120 Hz using a Flock of Birds magnetic system (Ascension Technology). If the cursor hit the target, a pleasant tone was played. Subjects also received numerical feedback indicating their speed, and were told that this score indicated solely their speed. Subjects were told that the object of the task was to hit the target, while maintaining a quick speed; they were not required to stop at the target. Subjects took 118 ms to reach the target extent, on average.

Our primary concern was to expose the subjects to a visuomotor perturbation and test whether behavior after adaptation exhibited steady-state errors. To understand the reason for these errors, we then probed how subjects learned from error by following the perturbation with error-clamp trials in which we de-coupled visual feedback from reach angles, controlling for error on each trial.

All subjects first completed 40 baseline trials, where cursor position was veridical, followed by a short break. They then completed 20 additional baseline trials, followed by 100 trials of training in which the cursor was rotated 30

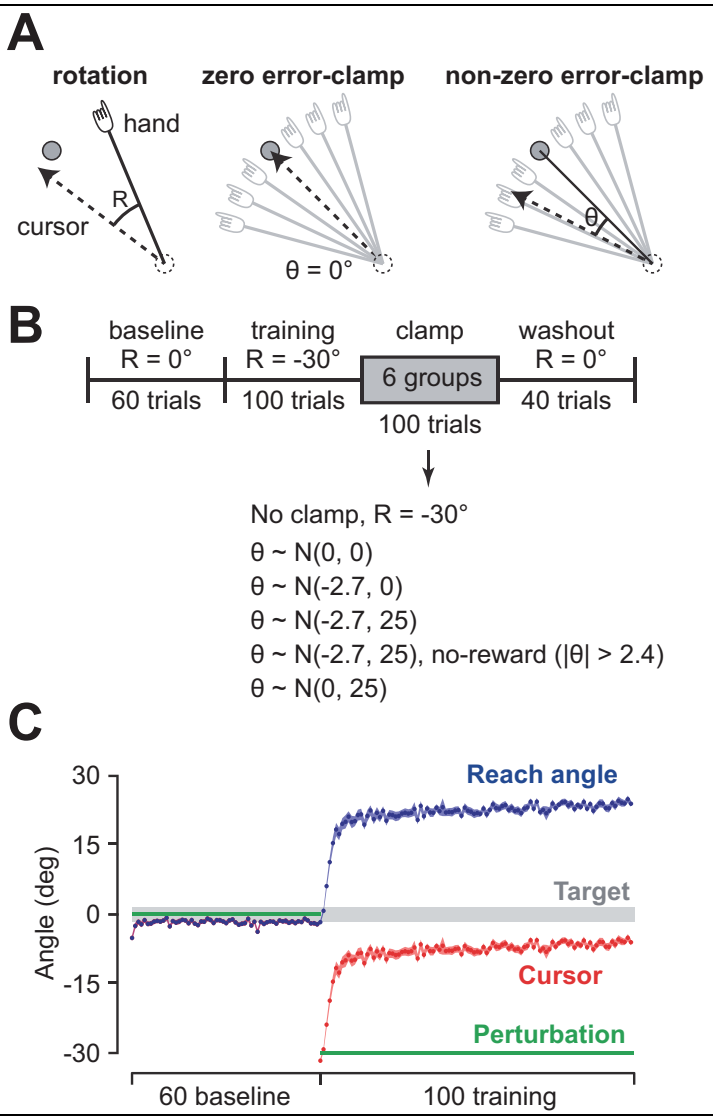

Figure 3.1**:** Experimental paradigm. **A**. Subjects made 8 cm reaching movements to a target at 135°. Visual feedback was perturbed by a rotation (left) or clamp (middle and right). In clamp trials, the cursor (dashed) moved directly to the target (middle) or to an endpoint, independent of the reach angle (solid). **B**. Subjects completed 60 trials of baseline training with veridical visual feedback followed by 100 trials of a 30 degree visuomotor rotation. Then, subjects were exposed to 100 clamp trials, where visual feedback was presented along a line at a fixed angle in each trial. **C.** Each group received a different clamp angle (θ) distribution. We denote these distributions by N(x, y), indicating the feedback was normally distributed with mean x and variance y. A control group (**no-clamp**) received additional rotation trials instead of clamps. One group was always clamped to the target (**N(0, 0)**). Three groups received feedback with a distribution that was similar in mean and/or variance to their movements during the training period (**N(0, 25)**, **N(-2.7, 0)**, **N(-2.7, 25)**). A last group received feedback with similar mean and variance, but trajectories to the target ($|\theta| < 2.4°$) resampled so no trials were rewarded (**N(-2.7, 25),no-reward**).

degrees CCW. Each group then completed a group-specific pattern of clamp-trials in which we controlled the trial-to-trial distribution of visual feedback, followed by 40 trials of washout in which veridical feedback was provided Figure 3.1B).

We tested 6 groups of subjects (Figure 3.1C). In the first group (no-clamp), subjects received 100 additional perturbation trials without an error-clamp block. In the next five conditions, subjects were presented with error-clamp trials, where cursor position was decoupled from hand position, but we modulated the distribution of cursor feedback. We use the notation N(x, y) to indicate that subjects received visual feedback drawn from a normal distribution with mean x and variance y. The N(0, 0) group (n = 7) received traditional error clamp trials, where the cursor moved towards the center of the target on every trial. That is, the cursor feedback had a distribution with mean zero and variance zero. In pilot experiments, after 100 trials of training in a 30°CCW rotation, subjects were receiving feedback with a distribution with mean -2.7° and standard deviation 5° (variance 25°). Accordingly, the N(-2.7, 25) group (n = 10) received cursor feedback drawn from this distribution. These parameters were similar to the mean and variance of the distribution of visual feedback these subjects received at the end of the training block. To examine the role of reward, we considered a condition where the distribution of feedback resembled that during training, but without any successful trials. In the N(-2.7, 25), no reward group (n = 8), any

trials where the cursor would end in the target were resampled from the distribution. To dissociate the roles of mean and variance of the feedback distribution on behavior, in two groups we used a distribution where one of these parameters was similar to that of subjects at the end of training, but the other changed. The N(0, 25) group (n = 10) received visual feedback with a distribution with mean zero and variance 25 deg^2, a variance similar to the variance of subjects' movements at the end of the training block, but with a different mean. The N(-2.7, 0) group (n = 15) received feedback with a bias (-2.7°) but no variance.

State-space model

We fit a two-state model to the behavior of subjects during the training block, and assessed to what degree the model fit could predict behavior during the clamp and washout block of the experiment (Smith et al., 2006). In this model, the perturbation was estimated via two internal states, one of which learns quickly from prediction error, but also reverts to baseline quickly, called the "fast" state, and a second which learns slowly and reverts to baseline slowly, termed the "slow" state. On each trial, the state estimate $x^{(n)}$ is updated from the error $e^{(n)}$ as follows:

$$\begin{aligned} x^{(n)} &= x_f^{(n)} + x_s^{(n)} \\ x_f^{(n+1)} &= A_f x_f^{(n)} + B_f e^{(n)} \\ x_s^{(n+1)} &= A_s x_s^{(n)} + B_s e^{(n)} \end{aligned} \qquad \text{Eq. 3.1}$$

Where, in Eq. 3.1, x_f and x_s are the fast and slow states, respectively. The learning rates are $0<B_s<B_f<1$ and the retention rates are $0<A_f<A_s<1$. For each subject, we fit the learning and retention rates of the model to the behavior during the baseline and training block, minimizing the least squared error between the data and model output. Then we fixed the parameters and, using the feedback provided in the clamp block, predicted the reach angles during the clamp and washout blocks for each subject. To evaluate the ability of the model to accurately predict behavior, we computed the variance accounted for (VAF) by the model in the fit and prediction periods as well as the square root of the mean squared error (RMSE) of the model predictions. We also computed the predicted trial-to-trial retention at the end of training. To do so, we calculated the retention in the state estimate at the end of training if a trial without error were presented: $\left(A_f x_f^{(n)} + A_s x_s^{(n)}\right) \Big/ x^{(n)}$.

Data Analysis

Data was analyzed offline using custom routines written in MATLAB (R2013a, The Mathworks). The angle of the hand at the target extent, relative to the target direction, was used to assess direction of the movement. Movements in the wrong direction (> 120° from the goal direction) were eliminated (0.04% of trials). All measures are reported as mean±SEM.

Statistical analyses were conducted using MATLAB (R2013a, Mathworks) or SPSS (V21, IBM). We used Student's t-test (paired when appropriate) to compare performance. Because the VAF by the state space model is bounded to the interval [0, 1] and not normally distributed, Mann Whitney U and Kruskal Wallis tests was used to compare the VAF across groups.

Exploration

Prior work has shown that when subjects receive limited feedback about their movements, substantial trial-to-trial variability in behavior can be observed (Izawa and Shadmehr, 2011). Recent work has further suggested that subjects can make use of variability to arrive at a successful solution when a perturbation is applied (Wu et al., 2014). We sought to identify exploratory behavior in clamp trials, where subjects might increase their trial to trial variability in the face of altered feedback. Because subjects differ in their behavior, we compared the variability of each subject's behavior in clamp trials to the variability of their behavior in the training block. We first calculated the standard deviation (SD) of the hand direction in an 11 trial (current ± 5 trials) sliding window across the experiment to find the variability of behavior around each trial. By finding the mean and SD of this distribution in the last half of the training block, we could evaluate the typical variability for each subject. Then, we identified windows in the clamp block about which subjects had a high variability, more than 2 SDs from the mean for that subject. For a given trial, if more than 80% of the 15

nearby trials (current ± 7 trials) were considered to have a high variability, we label the point as representative of exploration. This second criteria was used to prevent a single, potentially erroneous movement from causing several trials to be labeled as exploratory, by requiring a sustained increase in the trial to trial variability in behavior.

Section 3.3: Results

We asked six groups of volunteers to make 8 cm reaching movements. In the baseline block the motion of the cursor was an identity transformation of the motion of the hand, and the participants reached to place the cursor in the target. In the training block the motion of the cursor (red dots, Figure 3.1C) was a -30° (CCW) rotation of the motion of the hand (blue dots, Figure 3.1C). Participants learned to alter their reach angle by an average of 25.5±0.5° (last 20 trials of the training block compared to the last 20 trials of the baseline block, across all groups), resulting in 85% compensation for the perturbation, producing a residual error of -6.3±0.5°, a value significantly different than zero (t-test, $t(56) = -12.0$, $p = 4E-17$). However, the target radius was only 1.8° – why did participants exhibit a residual error, on average missing the target?

A state-space model of adaptation provides one account of this behavior (Smith et al., 2006; Kording et al., 2007). According to this model, in the training block the subject notes that as they generate a motor command, the cursor does not go where they had expected. The difference between the expected sensory

consequence and the observed consequence is a prediction error, inducing learning in a forward model relating motor commands to their expected sensory consequences. State space models assume that this internal model is parameterized via a set of states, and that these states learn from prediction error. Critically, these states also partially revert toward their baseline state following each trial. Consequently, subjects will converge on a non-zero error at the end of training when incomplete retention from one trial to the next, pulling behavior toward the baseline, balances new learning from the residual error. A two-state model fitted to the data for all subjects in the baseline and training block estimated that trial-to-trial retention at the end of training was 0.95±0.01. That is, the model predicted that without errors, motor output will decay at the rate of about 5% per trial at the end of training. We used zero-error clamp trials (trials in which motion of the cursor always hit the target, regardless of subjects' actions) to quantitatively test this prediction. We then used non-zero error clamps, in which the errors were drawn from various distributions similar to subjects' behavior at asymptote. We hypothesized that presenting errors that were decorrelated from subjects' actions might trigger subjects to abandon their steady-state behavior (Vaswani and Shadmehr, 2013) and lead them to overcome this residual error.

State-space model accounts for persistent residual error

In the No-clamp group the visuomotor rotation was maintained for 200 trials. With this group we wished to answer two questions: would performance exhibit residual error even with this extended amount of training, and if so, would the state-space model from the first 100 training trials predict performance in the second 100 training trials.

Performance of a typical subject is shown in the left column of Figure 3.2A and the group data are summarized in the right column of the same figure. With extended training the reach angles changed, resulting in partial compensation for the perturbation and placement of the cursor near the target. Indeed, by the end of the extended training the cursor position exhibited a non-zero error (last 20 trials, error of -4.2±0.8°, $t(7) = -5.2$, $p = 0.001$). We fit a state-space model to each participant's data in the baseline and first 100 training trials, and then fixed the model parameters and used it to predict performance in the remaining 100 training trials, as well as in the washout trials (black dots, Figure 3.2A). The model fit the baseline and training data well, accounting for 87±2% of the variance, with a root mean squared error (RMSE) of 4.9±0.3°. The model estimated that trial-to-trial retention after 100 trials of training was 0.95±0.02. Overall, the model predicted behavior in the extended training and washout block of the experiment quite accurately (variance accounted for, VAF 86±2%; RMSE 4.5±0.2°).

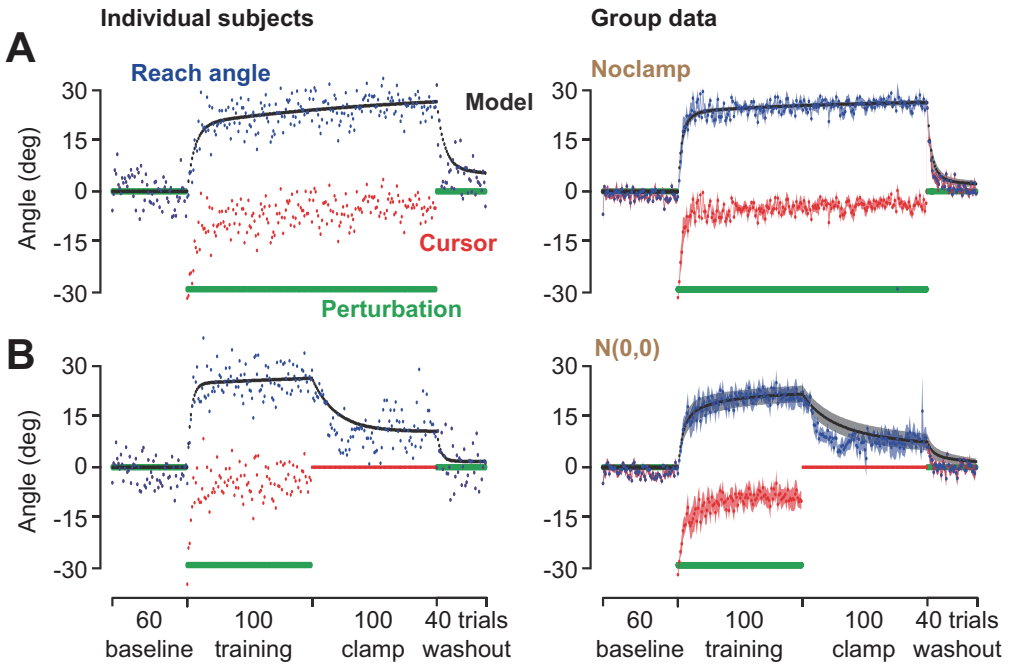

Figure 3.2: Behavior with extended training and in clamp trials to zero. Behavior of typical subjects (left) and groups (mean±SEM across subjects, right). Green lines indicate the rotation angle (R). Red points indicate cursor feedback. Blue points indicate the reach angle. A state space model was fit to the baseline and training data for each subject. The parameters from that fit were fixed and used to predict behavior in the clamp and washout periods. Black lines indicate the predictions of the model (mean±SEM across subjects).

Therefore, in extended training participants continued to exhibit significant residual errors, and these errors could be captured by a model in which there was a small but significant trial-to-trial reversion towards baseline.

Reversion to baseline incompletely predicted by a state-space model

If the residual error is due to trial-to-trial reversion to baseline, then in the absence of error the reach angles should decay toward baseline precisely as predicted by the model. To test for this, we presented a second group of subjects with traditional error-clamp trials where, regardless of the reach angle, the cursor moved towards the center of the target. We label this kind of error-clamp

as having a mean zero, variance zero distribution associated with error, N(0, 0). Performance of a typical subject that experienced this condition is illustrated in the left column of Figure 3.2B. In the training block the reach angles changed (blue dots), resulting in partial compensation for the perturbation and placement of the cursor near the target. Importantly, by near the end of training the cursor positions exhibited both a residual mean error (-3.4°), and variance (11.6 deg^2). At the onset of the N(0, 0) error-clamp, the cursor position became decoupled from reach angles. Consequently, the reach angles exhibited a change toward baseline.

We fit a state-space model to the data in the baseline and training trials, and then fixed the model parameters and used it to predict performance in the error-clamp. The model accounted for 91% of the variance in the reach angle of this subject over the range of trials fitted, with an RMSE of 4.2°, and produced an estimate of trial-to-trial retention of 0.96±0.02 across subjects. The model predicted that at the onset of the error-clamp trials the reach angles should gradually revert toward baseline, and that this reversion would not be complete by the end of the block, resulting in a non-zero bias (black dots, left column, Figure 3.2B). We observed both of these predictions in the data collected from this subject. However, it appeared that the actual decay in reach angles was somewhat faster than predicted (blue vs. the black dots, left column, Figure 3.2B).

To analyze the group data, we again fit a state-space model to the baseline and training data of each subject, and then computed the model's predictions for the error-clamp and washout blocks. By the end of the training block the reach angles exhibited residual error of -9.1±2.6°, an amount significantly different than zero ($t(6) = 3.5$, $p = 0.01$). The model fit the baseline and training data well (VAF, 84±4%; RMSE, 4.2±0.1°), and then predicted a decay in reach angles during the N(0, 0) error-clamp, producing a predicted bias of 7.8±2.6° at the end of that block (black line, right column, Figure 3.2B). Indeed, in the measured data we found that reach angles in the error-clamp block declined (end of training vs. end of clamp, paired t-test, $t(6) = 6.17$, $p < 0.001$), resulting in a bias of 8.4±2.9° at the end of the block. This bias at end of the error-clamp block was a fraction (32±15%) of the reach angles achieved during the training block, a fraction that was similar to that observed in our previous work in a force field learning task (26±5%, Figure 2.2). Therefore, exposure to the N(0, 0) error-clamp resulted in a reversion of the reach angles toward baseline, with an endpoint that was well predicted by the model for a clamp trial block of this duration.

However, the model predictions and measurements differed in one aspect. Upon introduction of the error-clamp trials the reach angles changed more rapidly than was predicted by the model (paired t-test, first 20 trials of clamps, $t(6) = -2.5$, $p < 0.05$). Overall, the model did a modest job predicting the data in the clamp and washout blocks (VAF, 42±8%; RMSE, 7.0±0.9°).

In summary, the results of the N(0, 0) group illustrated that in the error-clamp block the motor output decayed toward baseline with an endpoint that was well predicted by the model, but with a decay rate that was significantly faster than predicted. The inability of the model to fully predict the data is important because it puts in doubt the applicability of the state-space model, and the interpretation that it provides for residual errors at the end of training. However, another possibility is that the state-space model does provide an accurate description of one kind of learning that occurs in adaptation tasks, but that this specific form of error-clamp trial is not an innocuous probe, instead transitioning behavior to an alternative learning process evidenced by, in this case, a faster reversion to baseline.

Exploratory behavior in non-zero error clamps

We next considered an error-clamp in which instead of zero error, subjects were presented with a small, constant, non-zero error (N(-2.7, 0) group). Importantly, the error in the error-clamp block was smaller than the participants' residual error in the training block. Our hypothesis was that a sudden de-correlation between actions and errors in the setting of a residual error might trigger processes sensitive to residual target error and overcome the steady state reached by adaptation. As we will show, the state space model predicted partial reversion to baseline. Behavior, however, was dramatically different than predicted.

Following the training block, volunteers in this group were exposed to a distribution of visual feedback in which the mean error was -2.7°, with zero variance. Performance of a typical subject in the N(-2.7, 0) group is shown in the left column of Figure 3.3A. By the end of the training block this participant's reach angles produced a residual error of -7.2°. Note that the magnitude of this error was larger than the errors presented in the error-clamp block. As a consequence, the state-space model predicted that the reach angles would revert partially towards baseline (black dots, Figure 3.3A). However, the participant's behavior was qualitatively different than predicted: reach angles did not monotonically revert towards baseline, but instead varied dramatically from trial to trial. It appeared that the participant was searching for a reach angle that would place the cursor in the target.

The group data are shown in the right column of Figure 3.3A. By the end of the training block, movements exhibited a residual error of -5.8±0.6°. Remarkably, when presented with a smaller error in the error-clamp block, rather than maintaining their performance or reverting towards baseline, the subjects on average increased their reach angle (end of training vs. end of clamp, $t(14) = 2.6$, $p = 0.02$), attempting to close the small but persistent error. By the end of the clamp block, they reached on average 32.1±3.4° from the target. This behavior in the N(-2.7°,0°) error-clamp trials was interesting because it was quite different than expected from the standpoint of the behavior in the training trials.

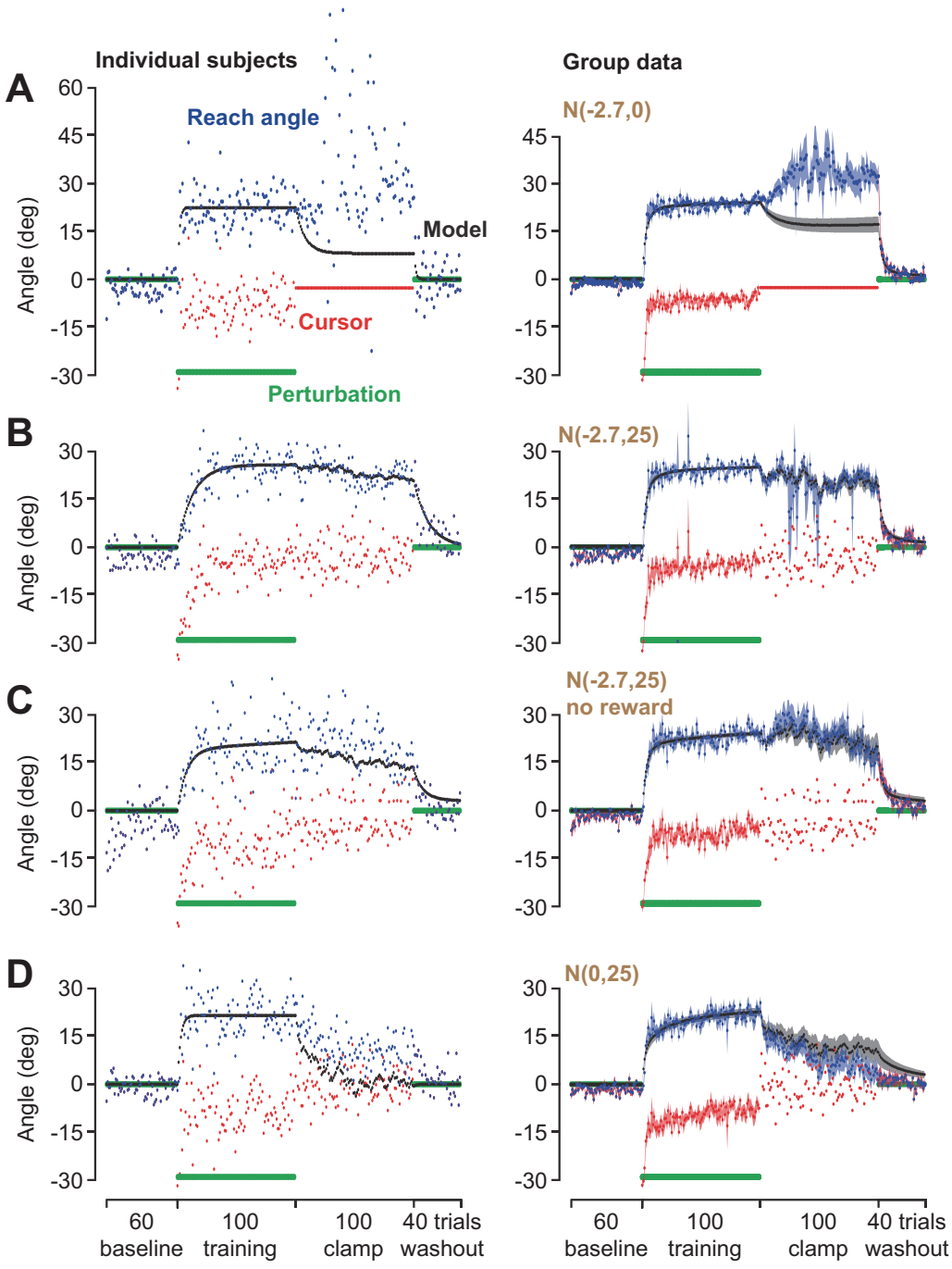

Figure 3.3: Behavior in clamp trials with non-zero mean and variability. Behavior of typical subjects (left) and groups (mean±SEM across subjects, right). Green lines indicate the rotation angle (R). Red points indicate cursor feedback. Blue points indicate the reach angle. A state space model was fit to the baseline and training data for each subject. The parameters from that fit were fixed and used to predict behavior in the clamp and washout periods. Black lines indicate the predictions of the model (mean±SEM across subjects).

Subjects maintained a persistent error of -5.8°, on average, in the training trials, but attempted to compensate for a smaller, -2.7° error in the clamp trials.

We fit the state-space model to each participant's data in the training trials and then used the model to predict behavior in the error-clamp and washout trials. The model estimated a trial-to-trial retention of 0.95±0.01 and predicted that reach angles should decrease in clamp trials because of the reduction in mean error size. However, this did not occur, as on average reach angles increased (Figure 3.3A). Indeed, the model was a poor predictor of behavior in the error-clamp block (fit period VAF, 83±2%, RMSE, 5.4±0.3°; prediction period, VAF, 51±4%, RMSE, 17.1±2.5°). Behavior in this group appears to reflect a transition towards using an alternative learning mechanism in order to close the constant residual error. As we will show later, instead of learning from error, soon after start of the error-clamp block some subjects in this group behaved in a way that suggested exploration.

Learning from error in error clamps with non-zero mean and non-zero variance

To further ascertain what it was about the non-zero error clamp that led to the exploratory behavior, we tested a new N(-2.7, 25) group, in which subjects were presented with the same constant, small bias, on average, but also with a variance similar to the subject's own performance in the training block. In previous work we have shown that giving an error clamp with variance similar

to that seen during initial adaptation prevents the reversion to baseline predicted by a state space model (Vaswani and Shadmehr, 2013). Thus if the exploratory behavior seen in the N(-2.7, 0) group was the result of detecting the change to non-zero clamp with zero variance, perhaps detection would be more difficult with a more naturalistic variance of errors in the non-zero error clamp. Performance of a typical subject in the N(-2.7, 25) group is shown in the left column of Figure 3.3B. At the end of the training block, the subject had a mean residual error of -4.1° with a variance of 12.0 deg^2 in the errors experienced. We fit the model to the data in the baseline and training blocks of this subject, and found that it well predicted that subject's performance in the ensuring error-clamp and washout blocks (left column of Figure 3.3B, fit period, VAF, 92%, RMSE, 4.3°; prediction period, VAF, 84%, RMSE, 3.8°).

This excellent prediction ability was true in the group data as well. In the N(-2.7, 25) group data, by the end of the training block the movements exhibited a residual error of mean -4.9±0.5° and variance of 19.3±1.5 deg^2. When we fit the model to the training trials for each subject in the N(-2.7, 25) group and then used the model to predict behavior in the remaining trials, we found that in the error-clamp and washout blocks the measured behavior was well predicted by the model (Figure 3.3B) (fit period VAF 85±3%, RMSE, 5.6±0.6°; prediction period VAF, 67±8%, RMSE, 9.4±2.8°). Indeed, the model did a better job of predicting behavior in the N(-2.7, 25) error-clamp block as compared to N(-2.7, 0) group

(Mann Whitney U-test, $U(9,15) = 26$, $p < 0.02$), as well as better than the N(0, 0) group (Mann Whitney U-test, $U(7,9) = 9$, $p < 0.02$). That is, it appeared that when the feedback distribution in the error-clamp block had a non-zero mean and variance, a state space model that accounted for a subject's behavior during training could also account for behavior in clamp trials. This was the case even when the distribution of feedback had a reduced bias as compared to the training block. The exploratory behavior seen in the N(-2.7, 0) group, in which there was a small non-zero error with zero variance, was not seen in the setting of normal variance. This suggests that error-based learning described by the state-space model is the default learning system during adaptation unless a change in the statistics of movements is detected.

The state-space model assumes that changes in behavior are driven by prediction errors in sensory outcomes of motor commands. We have previously shown that adaptation based on sensory predictions errors will proceed at the expense of decreasing task success (hitting the target goal) (Mazzoni and Krakauer, 2006). This led us to conclude that cerebellar-based sensory-prediction error-dependent adaptation is indifferent to reward (Krakauer et al., 2006; Krakauer and Mazzoni, 2011; Izawa et al., 2012; Haith and Krakauer, 2013). If this is true, then it predicts that behavior predicted by a state-space model in a non-zero error clamp should proceed in the absence of task success (hitting the target). To test this prediction, we considered an error-clamp condition where

distribution of feedback had both a variance and a small bias, but none of the trials produced a rewarding outcome during the error-clamp block (Figure 3.3C). In this N(-2.7, 25) no-reward group, any randomly generated cursor position in the error-clamp block that would land in the target was discarded and resampled from the N(-2.7, 25) distribution. For each subject we fit the model to the data in the baseline and training blocks and then used the model to predict behavior in the remaining trials. We found that despite the absence of rewarding trials, performance in the error-clamp was again well predicted (fit period VAF, 83±2%, RMSE, 5.4±0.4°; prediction period VAF, 70±5%, RMSE, 7.1±0.5°). That is, even without task success, error-based learning in the training trials was largely maintained in the N(-2.7,2.5)no-reward group as in the N(-2.7,2.5) group.

Finally, we considered a scenario in which the feedback had a variance similar to subjects' own movements, but no bias: N(0, 25) group. In this group, by the end of the training block the movements exhibited a bias of 8.1±1.8° and variance of 23.4±3.3 deg^2. We fit the model to the baseline and training trials for each subject and then used it predict behavior in the error-clamp block. The model predicted decay in the error-clamp block, but the observed decay was again faster (Figure 3.3D). Indeed, the model did a poor job of predicting behavior in the error-clamp and washout blocks (fit period VAF, 78±3%, RMSE, 5.8±0.3°; prediction period VAF, 33±5%, RMSE, 10.6±1.0°). As in the N(0, 0) group, the measured behavior exhibited a decay that was faster than predicted

by the model, suggesting a change in policy also occurred when the bias of the distribution of feedback was altered in clamp trials.

Overall, the results suggest that when sensory feedback is decoupled from the motor commands of the subject, a condition that is met in all error-clamp trials, the error-dependent learning policy in the training trials is largely maintained if the distribution of errors in the error-clamp trials (including mean and variance) are similar to subjects' own patterns in the preceding training trials. To statistically test this idea, we used a non-parametric ANOVA to test the effect of error distribution on the ability of the state space model to account for behavior in the clamp and washout blocks. There was a significant difference in the prediction period VAF across groups (Kruskal Wallis, $\chi^2(5) = 32.4$, $p < 0.00001$). We compared the prediction period VAF for each group to the VAF in the no-clamp group. Post hoc tests of mean rank revealed no significant reduction in the ability of the model to account for behavior in the N(-2.7, 25) and N(-2.7, 25) no-reward error-clamp blocks as compared to the no-clamp group. That implies that the learning policy exhibited in the training trials was largely maintained in the N(-2.7, 25) and N(-2.7, 25) no-reward error-clamp blocks. However, the prediction period VAF in the N(0, 0), N(-2.7, 0), and N(0, 25) groups was significantly different than that in the extended training group. That is, when the error-clamp block presented an error distribution that was similar to errors during training, behavior appeared consistent with the training trials,

exhibiting learning from error and partial reversion to baseline. When the error-clamp block presented an error distribution that was different from the training trials, the learning policy changed, producing behavior that appeared inconsistent with the training trials.

Quantifying exploration

As mentioned above, participants in the N(-2.7, 0) group showed behavior that was quite different than what would be expected by a state-space model: in clamp trials, they dramatically increased their trial-to-trial variability (Figure 3.4A). This increased variability appeared to us to be a form of exploration. We attempted to quantify this behavior.

Because subjects can have very different variability in their reach direction across movements, we defined exploration as a sustained increase (persistently more than 2 SD above the mean) in the trial-to-trial variability of a subject's movements as compared to the typical trial-to-trial variability observed in the training block for that subject (Figure 3.4A, lower panels). This metric appeared to successfully capture the trials in which subjects were exploring. Subjects in the N(-2.7, 0) group explored an average of 19±6.5 trials. This group included 3 subjects who explored for more than half of the 100 clamp trials. No subjects in the No-clamp group and only 1 subject in the N(0, 0) group showed any evidence of exploration (Figure 3.4B).

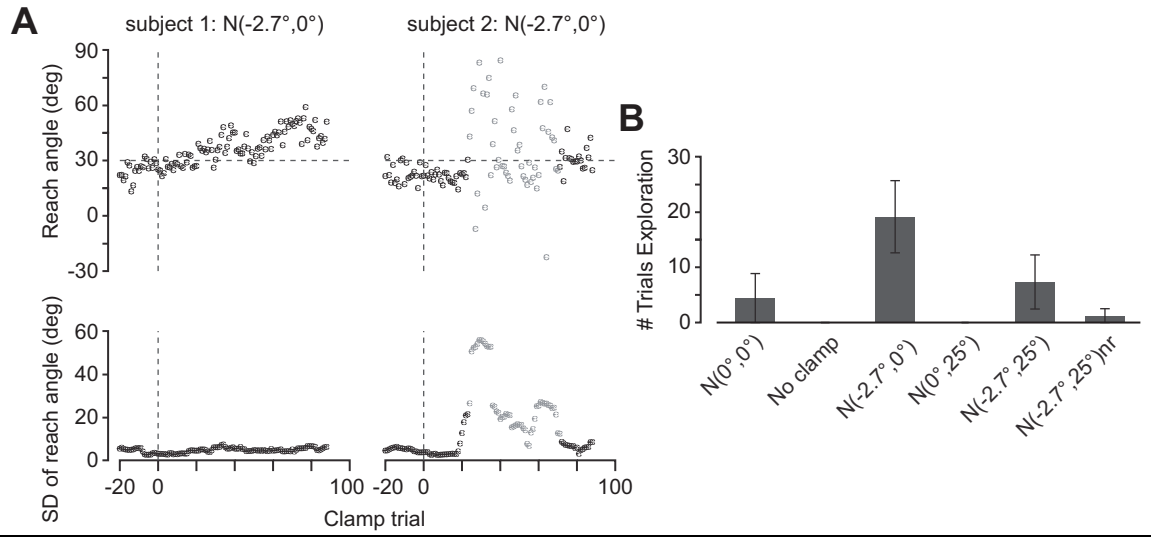

Figure 3.4: Exploration in clamp trials. **A.** In the N(-2.7, 0) group, some subjects tried to close the error gradually (top left), while some dramatically increased the trial to trial variability of their movements (top right). We identified these trials in which subjects were exploring (grey) by comparing the variance of subjects' movements to their variance during the training period (bottom). **B**. The number of trials identified as exploration (mean±SEM across subjects) in the 6 groups. Subjects in the N(-2.7, 0) explored more than any other group, including 3 subjects who explore for more than 50 trials.

Why did subjects change their behavior so dramatically from that observed in training? During the training block, errors had a non-zero mean as well as a non-zero variance. In the N(-2.7, 0) error-clamp block, the cursor was placed 2.7° away from the center of the target, regardless of the movement produced by the subjects. Therefore, the onset of the N(-2.7, 0) error clamp block produced four kinds of change in the feedback: 1) a reduction in the mean of the distribution associated with the position of the cursor; 2) a change in the variance of this distribution; 3) a change in the rate of success, in terms of accurately bringing the cursor to the target; and 4) a change in the correlation between reach angles and cursor position (in error-clamp blocks, the two become de-coupled). Perhaps in the N(-2.7, 0) group, the sudden change in the feedback from the

training block to the error-clamp block was instrumental in altering the learning policy. Using the number of exploration trials as our proxy for a change in the learning policy, we found that the probability of change in policy was largest for the N(-2.7, 0) group and smallest in the N(0, 25) and N(-2.7, 25) no-reward groups.

Section 3.4: Discussion

When movements are perturbed, resulting in errors, the motor system rapidly adapts its output to reduce these errors (Shadmehr and Brashers-Krug, 1997; Krakauer et al., 1999, 2000; Thoroughman and Shadmehr, 2000). However, when a visual perturbation is presented repeatedly, subjects do not completely adapt. Instead, subjects sustain a residual error in their movements. Why do subjects select actions that result in these persistent residual errors?

Here, we demonstrated that even with extended training subjects sustained a residual error when learning to compensate for a visuomotor rotation. In the framework of a state space model, this residual error can be attributed to a balance between learning from errors and incomplete trial-to-trial retention. To test this idea, we used the state-space model to fit the data during training and then fixed model parameters and used it to predict behavior during conditions in which we controlled the error distribution, named error-clamp trials (Scheidt et al., 2000). We found that in error-clamp trials with a smaller error than the residual error and no variance, subjects changed their policy

dramatically and acted to close the error, exhibiting exploratory behavior while doing so. This change in policy was prevented if the distribution of feedback in the error-clamp block had a bias and variance similar to that in the training block (even when feedback precluded successful trials, and was unrelated to subjects' own motor commands). In these cases, the process that accounted for learning in the training block also accounted for behavior in the error-clamp block.

The results show that trial-to-trial retention and error-dependent learning processes that can describe behavior during adaptation can also largely predict behavior when errors are artificially imposed. However, the key novelty in our results is that this is only the case if the overall distribution of imposed errors is similar to that generated by subjects themselves. In contrast, when experiencing feedback that lacked realistic trial-to-trial bias and variability, subjects demonstrated changes in behavior that were qualitatively different from state-space model predictions. In particular, subjects tried to overcome a fixed imposed error that was in fact smaller than the mean residual error they converged on at asymptote in the training block, and exhibited a faster reversion to baseline when the distribution of imposed error had zero bias.

We, and others, have demonstrated that multiple learning processes likely operate in motor adaptation tasks, not just error-based learning (Diedrichsen et al., 2010; Huang et al., 2011; Izawa and Shadmehr, 2011). In particular, we have suggested that state-space models are good fits for prediction error-based

learning and that deviation from state-space predictions does not imply falsification of the model, but rather that additional learning processes are at play (Huang et al., 2011). Similarly, the current and previous experiments suggest that error clamps can both serve as read-outs for processes captured by state-space models and as a trigger for additional learning processes. Zero-error clamp trials, where feedback is constrained to zero and decoupled from subjects' behavior, have led to insights into learning processes that are well captured by state-space models (Hwang et al., 2006; Criscimagna-Hemminger and Shadmehr, 2008; Tanaka et al., 2009). An increasing body of evidence, however, has demonstrated that zero-error clamps are not innocuous but can induce subjects to actively change their learning policy (Pekny et al., 2011; Shmuelof et al., 2011; Vaswani and Shadmehr, 2013). Recent work suggests that the distribution and type of feedback available might determine the nature of the behavior elicited in clamp trials (Pekny et al., 2011; Shmuelof et al., 2011, 2012; Vaswani and Shadmehr, 2013).

Here, instead of presenting clamps where error was zero on every trial, we parameterized the distribution of errors. Using a state-space model, we fit the behavior during training and assayed the degree to which the learning policy during the training block predicted behavior during the clamp block. We found that when the distribution of feedback during the clamp block was similar in both mean and variance to the distribution in the training block, trial-to-trial

retention and learning from error continued unabated in the clamp trials. This remained true even when rewarding trials were excluded in the clamp block. On the other hand, in error clamps with zero error or zero variance, or zero error and natural variance, on average, reach angled decayed faster than predicted. Finally, when subjects were presented with a small, consistent error (N(-2.7, 0)), they attempted to close that error and exhibited exploratory behavior.

Thus the state-space model failed to predict behavior in the clamp trials in which the error distribution was very different than subject's performance in the training trials. The reason, we suggest, is that subjects are likely to alter their behavior when they detect a change in the distribution of errors (Shmuelof et al., 2011; Vaswani and Shadmehr, 2013). In Chapter 2 (Vaswani and Shadmehr, 2013), we presented subjects with a distribution of feedback that had a similar mean and variance to that during training. Here, we saw a qualitative change in the response to errors when either the mean or variance of the feedback distribution changed. Subjects, at times, dramatically increased the trial-to-trial variability in their movements in response to an error; this exploratory behavior could not be described by state-space models of learning. It is striking that, despite the availability of this alternate strategy, subjects did not use it to reduce residual error during the training block, nor when variability of feedback in clamp trials matched that seen at asymptote. We only observed exploratory

behavior when errors during the clamp block had a fixed mean and zero variance.

In a previous study, we showed that when feedback was altered to provide only binary feedback at the asymptote of the training block, there was a qualitative change in behavior in subsequent (zero-error) clamp trials (Shmuelof et al., 2012) compared to subjects who received both binary and full cursor feedback. Our explanation for this effect was that providing only binary feedback promoted engagement of an alternative learning system to maintain behavior at asymptote. The availability of vector error information from full cursor feedback must therefore have led to a suppression of this alternative learning mechanism. A similar selective suppression of a secondary learning process could explain the failure to overcome residual error at asymptote. Like the switch to binary error in prior studies (Shmuelof et al., 2012), a change in the distribution of feedback triggered engagement of a secondary learning process that was capable of reducing the residual error.

We suggest that this secondary learning process is reinforcement-based (Huang et al., 2011; Izawa and Shadmehr, 2011; Haith and Krakauer, 2013), which is consistent with the exploratory nature of the behavior. Furthermore, the lack of sensitivity to removal of reward when subjects received a realistic distribution of feedback in clamp trials suggests that this reinforcement-based component of learning plays little role in the behavior of the majority of the

groups we considered. Providing endpoint feedback alone appears to be another way to promote engagement of reinforcement-based learning (Izawa and Shadmehr, 2011). Notably, when subjects adapt given only endpoint feedback, they show more exploratory behavior early on and less asymptotic residual error (Taylor et al., 2014), further supporting the view that residual errors and reversion to baseline are not universal limitations to human motor learning but are a characteristic feature of error-based learning.

State space models may therefore capture a particular form of learning that dominates in normal adaptation tasks, at least during initial learning, when full cursor feedback is provided. This system leads to rapid reduction of errors, at the expense of a residual asymptotic error. What is the underlying learning mechanism that the state-space model captures so effectively? The error-driven component of the state-space model appears to relate to prediction errors driving an update to an internal forward model in the cerebellum (Mazzoni and Krakauer, 2006; Tseng et al., 2007; Taylor et al., 2010; Izawa and Shadmehr, 2011). The reversion to baseline could reflect passive decay in time of the parameters of this forward model. When experiments are conducted with longer inter-trial intervals, monkeys exhibit reduced learning and reduced complex-spike induced long term depression of Purkinje cells, indicating that cerebellar learning is affected by the passage of time (Yang and Lisberger, 2014). Passive decay may be advantageous because environmental processes themselves tend to dissipate

over time (Kording et al., 2007). Alternatively, reversion to baseline could reflect competition between a weakly and a strongly reinforced action (Shmuelof et al., 2012).

Conclusion

Residual errors in regular adaptation tasks reflect the operation of a single learning process (forward model-based, cerebellar-dependent) which, when provided with strong vector error, suppresses alternative forms of learning and is well-captured by a state-space model. This suppression can be overcome through changes in the distribution of feedback provided both during acquisition (Izawa and Shadmehr, 2011; Shmuelof et al., 2012) and during error clamp trials, as was done here. These changes in feedback lead to recruitment of an exploratory, possibly reinforcement-based, mechanism that is capable of reducing residual asymptotic errors. We conclude that error clamps can be "neutral" and capture the predicted retention behavior of state-space models when vector feedback is provided and when the distribution of errors matches those during acquisition. When these aspects of feedback are altered, other forms of learning are triggered that do not follow state-space model dynamics.

Chapter 4: Motor adaptation when maximizing reward is in conflict with minimizing errors

Section 4.1: Motivation

As the motor system executes a movement, it receives feedback from the sensory nervous system about the consequences of that movement. For example, when reaching with the hand toward an object, the brain receives visual and proprioceptive feedback about the state of the limb. The information provided by sensory feedback is used to build models of the body and environment, and select and refine subsequent actions (Shadmehr and Mussa-Ivaldi, 1994; Thoroughman and Shadmehr, 2000).

Several types of feedback and systems of learning are available when completing a motor task. Let us assume an action is generated with the aim of producing a desired sensory consequence – moving a cursor to a target to receive a point towards one's score, for example. The brain receives information about the errors in the sensory consequences of its action – how far the cursor missed the center of the target. The brain also receives information about task success – whether the cursor hit the target and if the movement was rewarded. Both of these sources of information have been demonstrated to play a role in the selection of subsequent actions, and are thought to involve distinct neural substrates. When feedback is only provided about success or failure of an action,

for example, subjects can use reward-based learning mechanisms adapt their movements to compensate for perturbations (Izawa and Shadmehr, 2011). Similarly, subjects alter their behavior to minimize the cost of error in the absence of feedback indicating task success (Körding and Wolpert, 2004).

Consider a simple adaptation experiment, where one must generate an action with the goal of producing feedback that is within a target (Figure 4.1). How should subjects select an action to accomplish that goal? A desired action produces a distribution of sensory consequences. Subjects have been shown to exhibit a supralinear cost of error, penalizing feedback that is further from the target (Körding and Wolpert, 2004). In this task, subjects could minimize the error-cost of feedback. Alternatively, they could use reward-based mechanisms (Izawa and Shadmehr, 2011) to select an action that maximizes task success. At baseline, these two mechanisms are congruous, and both indicate that the correct action that moves to the center of the target (Figure 4.1A).

In typical adaptation experiments, movements are perturbed, systematically altering the relationship between actions and their sensory consequences. In this case, subjects adapt, changing their actions such that the distribution of feedback continues to minimize the error cost and maximize task success (Figure 4.1B). Notably, these two goals remain congruous when a single, consistent perturbation is applied.

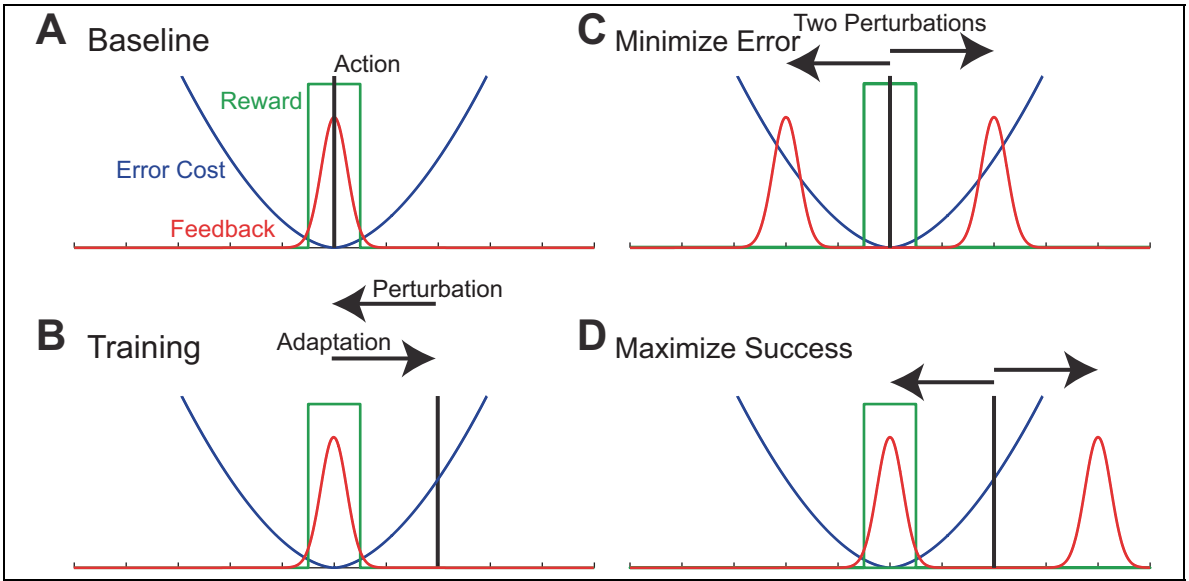

Figure 4.1: Intuition. Consider a simple one-dimensional task where one must produce an action (black), with the explicit goal of placing the resulting distribution of feedback (red) into the target (green). To do so, one can try to minimize an error cost (blue) – a function of the distance of the feedback from the center of the target. A, In typical adaptation paradigms, the action that minimizes an error cost is the same action that maximizes the rate of reward. B, When a perturbation is applied, changing the relationship between actions and their consequences (arrow), subjects learn to produce a new action. Typically, a single action will still both minimize the error cost and maximize the reward rate, as shown. C, When two perturbations are provided, these two goals are no longer congruous. Subjects can produce an action that minimizes the error cost of the resulting feedback, by producing an action that results in small errors when either perturbation is applied. D, Alternatively, subjects could produce an action that compensates for one of the perturbation, maximizing the reward but incurring a large error cost when the other perturbation is presented.

In contrast, several studies have used a simple paradigm to examine the ability of people to learn to compensate for two opposing perturbations simultaneously. In these experiments, actions that minimize the expected error cost of feedback are different from those that maximize the rate of success in the task. Subjects were typically presented with two, opposing perturbations (Figure 4.1C, D). That is, a single action produces a bimodal distribution of feedback. In this case, subjects have a choice. They can produce an action directly to the target, which results in feedback with small error cost. However, the sensory

consequences are never within the target, and so subjects will be unsuccessful. This behavior minimizes error, but produces no success (Figure 4.1C). Alternatively, subjects could produce an action that compensates for one of the perturbations. In this case, they will be successful on trials when that perturbation is presented. However, when the other perturbation is presented, they will experience large error costs. This behavior maximizes success, but produces large errors (Figure 4.1D). Prior studies have attempted to provide contextual cues to aid subjects in selecting the rewarding action on each trial. They demonstrate, however, that without strong spatial cues, subjects compensated for the mean perturbation and were largely unsuccessful – a behavior consistent with minimization of mean error (Karniel and Mussa-Ivaldi, 2002; Gupta and Ashe, 2007; Hirashima and Nozaki, 2012; Baldeo and Henriques, 2013; Howard et al., 2013). Other studies using cognitive cues to allow subjects to rapidly achieve task success by providing a strategy have also shown that subjects learn from error, at the expense of task success (Mazzoni and Krakauer, 2006; Taylor and Ivry, 2011).

We found this behavior puzzling. Subjects are able to compensate for the perturbations when each is presented alone (i.e. Figure 4.1B). In paradigms with two perturbations, subjects could achieve task success on 50% of trials if they learned to compensate for one perturbation and accept large errors when the other perturbation is presented. Why, then, do subjects adapt to the mean

perturbation, consistent with minimizing a cost of error, even though this behavior completely forsakes reward?

In the work in this chapter, we presented subjects with opposing perturbations, unpredictably, to place the error-based learning and reward-based learning systems in conflict. We tested several hypotheses to examine the conditions in which the motor system would select an action that results in task success, instead of minimizing error. Overall, we found that error-based learning provided a strong signal to the motor system that dominated behavior, even when it resulted in a reduction in task success. In the absence of prediction errors, the motor system maximized task success. Only when prediction errors were mild and the reward signal was strong did we observe a balance between the two learning systems. Otherwise, sensory prediction error dominated motor learning.

Section 4.2: Methods

Eighty-seven healthy, right-handed adults (Age 24.2±4.1 years old (mean±SD), 38 female), with no known neurological problems participated in this study. Participants were asked to hold the handle of a robotic manipulandum while making movements to a target. The hand and robotic arm were hidden from the subject by a screen, but visual feedback (3 mm cursor(s)) was projected onto a screen above the plane of movement of the arm. Visual

feedback was sometimes altered, as described below. The position, velocity, and force at the handle of the robotic arm were recorded at 200 Hz.

Subjects began their movements at a starting circle (1 cm diameter) at approximately the midline of the body. On each trial, a target (1 cm diameter) appeared 10 cm in front of the starting position and an auditory cue was played. Subjects were instructed to move the cursor rapidly and accurately through and past the target. Once subjects moved their hand more than 10 cm from the starting position, the cursor was hidden, feedback was provided, and a virtual cushion (resistive viscosity: 100 Ns/m) was presented to stop the arm.

Subjects received visual and auditory feedback based on the duration and accuracy of their movements. If subjects reached the target too quickly (< 150 ms), the target turned red and an error tone was played. If subjects reached the target too slowly (> 250 ms), the target turned blue and the error tone was played. If subjects moved with the appropriate duration, but missed the target, the target turned pink and the error tone was played. Finally, if subjects hit the target in the appropriate amount of time, the target 'exploded', a point was added to their score, and a pleasant tone was played. Participants were instructed to try to accumulate as many points as possible. After each trial the manipulandum returned the hand to the starting position. Short breaks of 1-2 minutes were provided periodically during the experiment.

Perturbations

Two types of perturbations were used in this experiment: forces applied to the hand, and visual perturbations applied to the cursor. In Experiments 1, 2, and 3.1, on some trials, forces were applied to the hand using a viscous curl field (Shadmehr and Mussa-Ivaldi, 1994). In these trials, the forces were described by $\mathbf{f} = kB\dot{\mathbf{x}}$, where $\mathbf{f}$ is the force vector, k is a constant indicating the field strength, $B = \begin{bmatrix} 0 & -15 \\ 15 & 0 \end{bmatrix}$ Ns/m is a viscosity matrix describing a counter clockwise (CCW) curl field, and $\dot{\mathbf{x}}$ is the hand velocity vector. In some trials, subjects were presented with an error-clamp (Scheidt et al., 2000), where the hand was constrained to a straight line from the start position to the target by a stiff, one dimensional spring (spring constant 6000 N/m, damping 250 Ns/m).

In Experiments 3.2, 4, and 5, visual perturbations were used. In these experiments, on some trials, the position of the cursor was rotated relative to the start position by a CCW rotation denoted by *R*. If the hand moved to the target, a rotation R = +15° indicates that the cursor would move 15° CCW of the target and miss it. For consistency, error-clamp trials were also used in these paradigms. In a visual error-clamp, the cursor was constrained to a straight line from the start position to the target, and only reflected the distance, not the angle, of the hand from the starting position.

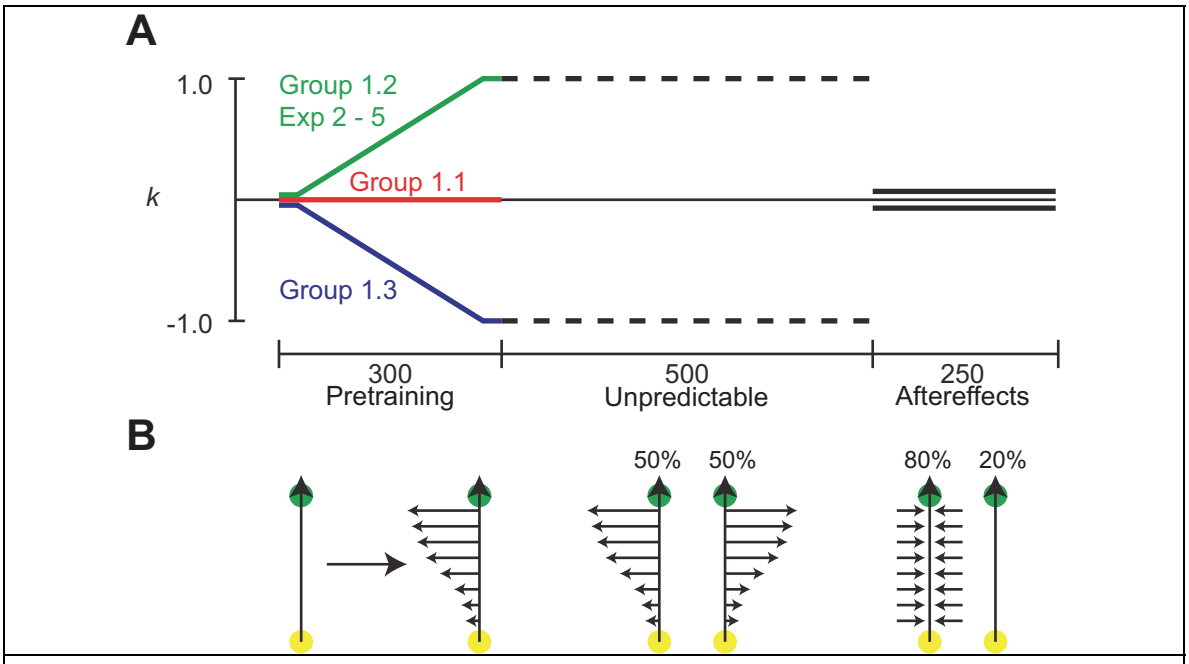

Figure 4.2: Experimental paradigm. A, Each experiment consisted of 3 epochs. In the pretraining period, subjects completed 300 trials where they received null training (Group 1.1), or were gradually presented with a rightward (k = -1, Group 1.3) or leftward (k = 1, Group 1.2 & Exp 2-5) perturbation. In the unpredictable period, leftward and rightward perturbations were presented pseudorandomly. In the aftereffect period, subjects received 80% error clamp and 20% unperturbed clamp trials. B, Example of perturbations for Group 1.2 for each epoch.

In each paradigm in this experiment, two perturbations of equal and opposite magnitude were used to generate a condition in which maximization of reward was at odds with minimization of error. In general, the paradigm for all experiments proceeded as follows (Figure 4.2A). First, subjects completed a 300 trial pretraining period. Then, they completed a 500 trial "unpredictable" period. In these trials, in some cases, a CCW perturbation to the left (k = +1, or R = +15) was provided (Figure 4.2B). In other cases, a clockwise (CW) perturbation to the right (k = -1, or R = -15) was provided. Because a single action cannot compensate for both perturbations, subjects had to make a choice. They could adopt an action that compensated for the leftward, or the rightward,

perturbation and forgo any reward associated with the other perturbation. Alternatively, they could adopt a policy that attempted to compensate for both perturbations, possibly minimizing the average performance or prediction error. Finally, subjects completed an aftereffect period, where unperturbed catch (k = 0, or R = 0) and error-clamp trials were presented (Figure 4.2B).

Experiment 1: Pretraining

In Experiment 1, our aim was to test whether learning the rewarding solution necessary to compensate for one perturbation would result in subjects maintaining that solution when presented with both perturbations. Three groups participated in this experiment. Participants in Group 1.1 (n = 8) served as the control group. These participants completed 300 trials of unperturbed movements (k = 0) in the pretraining period (Figure 4.3A). This was followed by 500 trials where a CCW curl field (k = 1) and a CW curl field (k = -1) were presented pseudorandomly, each 40% of the time (Figure 4.3B). The remaining 20% of trials were error-clamp trials to probe behavior. Finally, after this unpredictable period, subjects were presented with an aftereffect period, where 80% of trials were error-clamp trials and 20% of trials were unperturbed catch trials (k = 0).

Participants in Group 1.2 (n = 8) and Group 1.3 (n = 8) were gradually taught the solution to one of the perturbations during the pretraining period. In

Group 1.2, participants completed 25 unperturbed trials, followed by a series of 250 trials where the force field was increased linearly to full CCW strength, from k = 0 to k = 1. They then completed 25 movements with the full CCW field, k = 1, in the pretraining period. Subjects in Group 1.3 similarly were exposed to a ramp to a full CW (k = -1) field. After the pretraining period, subjects completed the unpredictable and aftereffect period described above.

Experiment 2: Repetition & reinforcement

In Experiment 2, our aim was to test whether repetition and reinforcement of the solution to compensate for one perturbation would result in maintenance of that solution in the unpredictable period. Subjects in Group 2.1 (n = 7) were tested on 5 consecutive days (Figure 4.4A). On the first day, they completed the same paradigm as Group 1.2, with gradual pretraining to a CCW perturbation, unpredictable presentation of CCW and CW perturbations, and an aftereffect period (Figure 4.4B). Then, on each of the second, third, and fourth days of training, they completed 500 trials where they were exposed to the CCW field (k = 1). Error-clamp trials were placed randomly with 20% probability during these reinforcement sessions. Finally, on the fifth day, they repeated the paradigm from the first day. We hypothesized that, with additional repetition and reinforcement of the actions that compensate for a CCW field, subjects may maintain that behavior in the unpredictable period on the fifth day.

Experiment 3: Reduced error feedback

On each trial of these experiments, subjects received three types of feedback. While making reaching movements, they received proprioceptive feedback about the position of the limb, continuous visual feedback in the form of the cursor, and feedback about the success of their movements, i.e. whether they hit the target or not. In the unpredictable period, minimizing the visual and/or proprioceptive errors would result in different behavior than maximizing the amount of success. Accordingly, in Experiment 3, our aim was to test if reducing the availability of proprioceptive and visual error would change the way actions are selected during the unpredictable period.

We reduced the visual error for subjects in Group 3.1 (n = 8). These participants completed the same paradigm as subjects in Group 1.2. However, we eliminated the midmovement visual feedback throughout the experiment. That is, after subject moved more than 2 cm away from the starting position, the cursor was hidden (Figure 4.5A, left). Only the endpoint of the movement – the point at which the subject reached 10 cm away from the start position – was shown.

We reduced the proprioceptive error for subjects in Group 3.2 (n = 8). Instead of receiving force field perturbations, which displace both the hand and, because feedback was veridical, the cursor, these subjects received only visual perturbations (Figure 4.5B, left). We rotated the cursor +15° or -15° in the

unpredictable period. Subjects in this group completed the same sequence of perturbations as subjects in Group 1.2, with visuomotor rotations instead of force fields. They first experienced a pertaining period, where they were gradually exposed to a CCW rotation of +15°. Next, they experienced the unpredictable period, where CCW and CW rotations were presented randomly. Finally, they completed an aftereffect period. Visual error-clamps were inserted on 20% of trials, as in Group 1.2.

Experiment 4: Reward in the absence of prediction error

In each of the preceding experiments, subjects were probabilistically exposed to two perturbations, one CCW and one CW, across trials. Because a single action can, at best, compensate for one perturbation, subjects will experience sensory prediction errors, differences between the expected and observed consequence of an action, and performance errors, differences between the desired and observed consequences of an action. Subjects also received reward prediction errors – a difference between the expected and desired/observed outcome of an action in terms of the target explosion.

In Experiment 4, we eliminated sensory prediction errors in the "unpredictable" period of the experiment, to test whether subjects make use of reward in their choice of an action. That is, we wondered if reward feedback –

the target explosion and points – mattered to subjects in this experiment. Two groups participated in this experiment.

Subjects in Group 4.1 (n = 8) experienced a visual perturbation, where a +15° rotation was gradually imposed upon a single cursor in the pretraining period. Then, in the "unpredictable" period, instead of presenting +15° and -15° rotations to a single cursor pseudorandomly across trials, two cursors were presented on every trial (Figure 4.6A, left) (Kasuga et al., 2013). One of these cursors was perturbed to the left, by +15° (the "left" cursor), and the other was perturbed by -15° (the "right" cursor). If either cursor hit the target in the appropriate time, subjects received reward feedback. Subjects were not informed about this success criterion and were told to "do their best." As in all experiments, on 20% of trials, visual error-clamp trials were provided, where a single cursor was constrained to a straight line to the target. Finally, in the aftereffect period, a single unperturbed cursor was again presented.

Subjects in Group 4.2 (n = 8) experienced the same paradigm as subjects in Group 4.1, with two cursors perturbed by +15°and -15° respectively in the "unpredictable" period (Figure 4.6B, left). However, to further test whether reward mattered to subjects in these experiments, reward feedback was provided to participants only if the -15° rotated cursor hit the target. That is, of the two sensory consequences, only the right cursor was rewarded. In the pretraining period, subjects were taught to compensate for a +15° rotation. When presented

with two cursors, initially the left cursor would hit the target, but not be rewarded. We wondered if subjects would learn to compensate for the -15° rotation of the right cursor in order to be successful, even though the left cursor was hitting the target.

Experiment 5: Sensory and reward prediction errors

Finally, in Experiment 5, we presented subjects with visual perturbations of two cursors simultaneously, as in Experiment 4, but we manipulated the sensory and reward prediction errors presented to the 3 groups.

Subjects in Group 5.1 (n = 8) completed the same paradigm as subjects in Group 4.1. However, in the unpredictable period, we imposed a mild, persistent sensory prediction error (SPE), by pseudorandomly coloring one cursor gray, instead of white (Figure 4.7A, left). As in Group 4.1, if either cursor hit the target, subjects were rewarded – the color of the cursor was irrelevant to the task. In this experiment, we wondered if subjects would maintain a rewarding solution, in the face of a mild sensory prediction error.

Subjects in Group 5.2 (n = 8) were also presented with a mild sensory prediction error by pseudorandom coloration of the cursors. However, in this group, we added a reward prediction error (RPE), so that, as in Exp 1-3, a single action could only produce reward on half of trials. In this group, subjects were rewarded only when the white cursor hit the target (Figure 4.7B, left). Because

the perturbation of the white cursor was pseudorandom on each trial, at best, a single action that compensated for only the leftward, or only the rightward, perturbation would only result in reward half the time.

Subjects in Group 5.3 (n = 8) were presented with only the reward prediction error. Participants in this group received two cursors on each trial in the unpredictable period. Both cursors were white, so that there were no persistent sensory prediction errors. However, only 1 cursor was rewarded, pseudorandomly (Figure 4.7C, left). As a result, a single action would only result in reward half the time, but the sensory consequences of that action were deterministic. That is, there was a persistent reward prediction error, but not a persistent sensory prediction error in this group.

Adaptation Index

We assessed performance in two main ways: the adaptation index of subjects on each trial and the success rate. The adaptation index is a measure of the degree of compensation for a perturbation of a full strength CCW (k = 1) field, or a full +15° rotation. Compensation in force field experiments was assessed using the force subjects applied in the error-clamp trials. The lateral force produced was regressed onto the ideal compensatory force profile for a CCW field. The measure indicates the degree to which subjects compensated for a velocity dependent force field. An adaptation index of +1 indicates subjects

fully compensated for a CCW force field; an adaptation index of 0 indicates no velocity dependent force was applied; and an adaptation index of -1 indicates subjects fully compensated for a CW force field. In visuomotor rotation experiments, the adaptation index was assessed using the reach angle of the hand. As in the force field task, compensation for a CCW (R = +15°) perturbation, indicated by a reach to -15°, was indicated by an adaptation index of +1. Compensation for a CW perturbation was indicated by a negative adaptation index.

We also assessed performance using the reward and hit rate. In tasks where force perturbations were applied, we computed the fraction of perturbed trials where the subjects received reward (reward rate). In visual perturbation tasks, we could compute the fraction of perturbed trials in which subjects received reward, or would have received reward if the other perturbation were present and rewarded (hit rate). Using the hit rate allows us to compare behavior across conditions where the rewarding criteria were manipulated in Experiment 5. Error-clamp trials were excluded from these analyses.

Statistics

Data analysis was conducted using custom routines written in MATLAB (R2013a) and SPSS (V22). Students t-test and repeated measures ANOVA were

used for statistical analysis, except where noted otherwise. All measures are reported as mean±SEM.

Section 4.3: Results

In the work in this chapter, we wondered what behaviors subjects would produce in a task where actions that minimize the cost of error were distinct from those which would maximize reward. We applied two opposing perturbations, either within or across trials, to create a condition where maximization of task success is distinct from minimization of error. In each experiments, subjects were presented with a choice. They could produce an action moving directly to the target, producing feedback with small error cost but forgoing success. Alternatively, they could produce an action that compensates for one of the perturbations, which will result in task success on 50% of trials but also produces occasional, large errors.

We wondered under what conditions subjects would choose to be successful and under what conditions they would forgo task success in order to minimize error. In each experiment, we asked subjects to make planar reaching movements to a target, perturbed their movements, and observed the actions they produced. We tested how training, reinforcement, feedback, and prediction errors affected subjects' behavior.

Experiment 1: Do subjects need to learn one solution first?

In Experiment 1.1, subjects completed 300 trials of pretraining in a null field, with no perturbation. We then presented 500 trials of unpredictable perturbations, followed by an aftereffect period. We assessed subjects' behavior in four ways.

First, we measured the errors experienced when each field was presented. If subjects were minimizing an error cost, we would expect errors of equal and opposite magnitude to each perturbation (Figure 4.1C). Alternatively, if subjects were attempting to achieve task success, we expect large errors to one perturbation, accompanied by small errors to the other (Figure 4.1D). Consistent with prior reports (Karniel and Mussa-Ivaldi, 2002; Gupta and Ashe, 2007; Hirashima and Nozaki, 2012; Baldeo and Henriques, 2013; Howard et al., 2013), we observed that subjects exhibited significant errors of similar magnitude to both perturbations even after 500 trials of exposure to the two perturbations (Figure 4.3C, right, last 50 trials of unpredictable period, 14.1±2.4° to CCW field, -16.0±1.6° to CW field).

We also measured subjects' behavior in error-clamp trials using an adaptation index. This measure is an assay for the "action," in Figure 4.1. If subjects were minimizing error, the adaptation index should be approximately zero (the mean of the two perturbations); alternatively, if subjects were trying to achieve task success, the adaptation index should be significantly biased, near

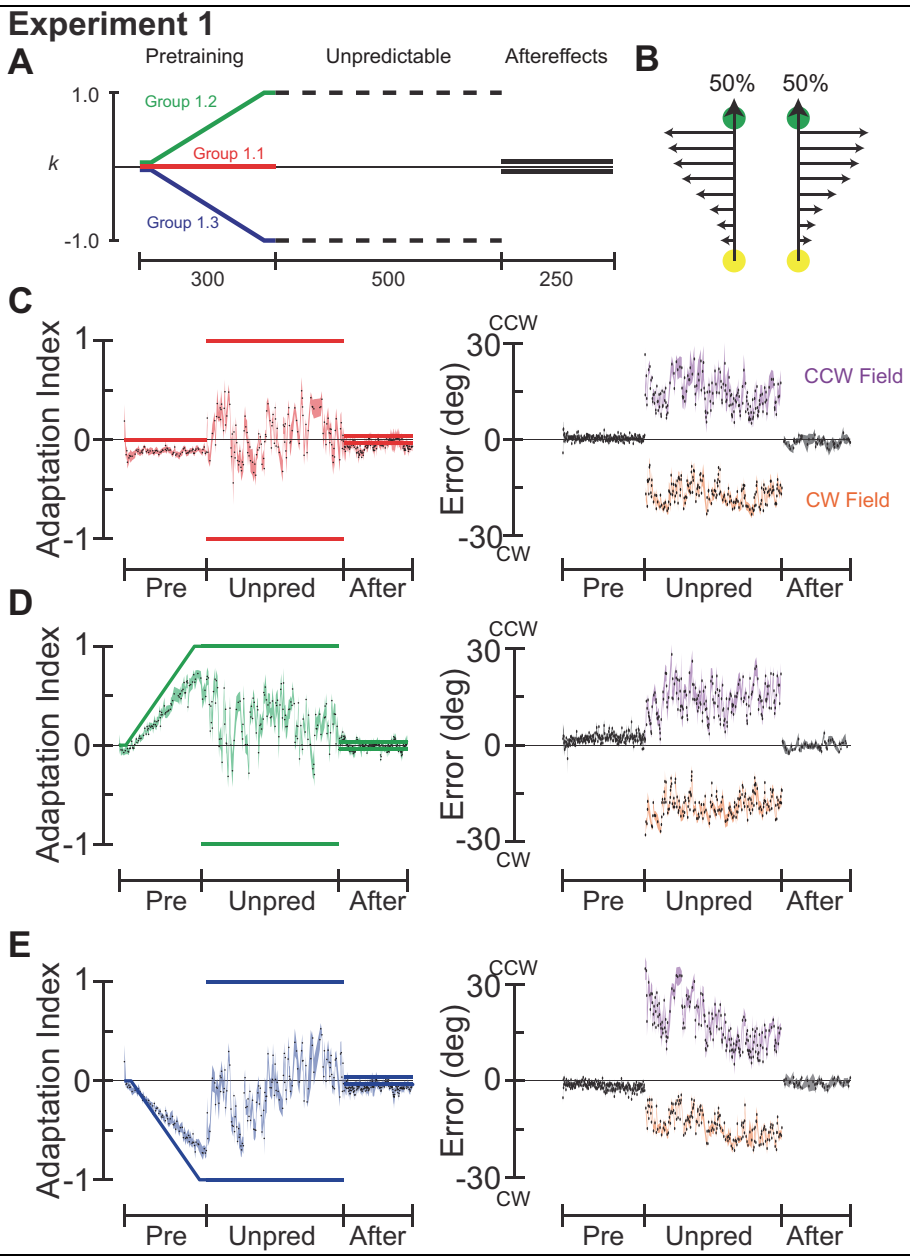

Figure 4.3: Experiment 1: Pretraining. A, In Experiment 1, subjects were gradually exposed to one perturbation, to test if they would maintain that solution when presented with both perturbations unpredictably. B, Leftward and rightward viscous curl fields were presented during the unpredictable period. C, Subjects in Group 1.1 did not produce actions that compensated for one perturbation, instead producing actions consistent with minimization of error. Behavior did not compensate for either perturbation by the end of the unpredictable period (left), and subjects experienced errors of similar magnitude when either perturbation was presented (right). D, Subjects in Group 1.2 and E, Subjects in Group 1.3 learned to compensate for CCW and CW perturbations during the pretraining period. They did not maintain that behavior during the unpredictable period (left). Initially, errors were smaller to the perturbation subjects learned (right), but by the end of the unpredictable period, subjects exhibited similar error to both perturbations. Subjects in all three groups did not have aftereffects. Data are mean±SEM across subjects.

one. We found that subjects produced behavior no different than zero (Figure 4.3C, left, last 50 trials of unpredictable period, adaptation index, -0.01±0.01, t-test vs. zero, $t(7) = -0.14$, $p = 0.9$).

Next, we assessed whether subjects were successful on the task. Subjects in this group were unsuccessful at the end of the unpredictable period (last 50 trials, success rate in non-channel trials, 1±1%). Finally, we assayed if subjects exhibited aftereffects, either in clamp or catch trials in the aftereffect period. Consistent with our other measures, subjects exhibited no aftereffects, in clamp trials (first 25 trials of aftereffects, adaptation index -0.06±0.06, t-test vs. zero, $t(7) = -1.05$, $p = 0.3$) or catch trials (-0.6±1.3°, t-test vs. zero, $t(7) = -0.45$, $p = 0.7$).

Therefore, overall we observed behavior consistent with minimization of the cost of error, and inconsistent with maximization of reward. However, we know that subjects are capable of learning to compensate for force field perturbations (Shadmehr and Mussa-Ivaldi, 1994) and can learn from feedback about success and failure (Izawa and Shadmehr, 2011). Why do they not do so?

Our first hypothesis was that subjects may not select and maintain a rewarding action because they are unable to learn to compensate for one perturbation in the presence of another. That is, there may be interference during the learning process. Compensation for a viscous force field is a complex behavior, and so the action that would compensate for one perturbation may not be evident to the motor system. To test this hypothesis, we allowed subjects in 2

groups to learn to compensate for a CCW or CW perturbation, prior to exposure to both fields unpredictably.

Subjects in Group 1.2 completed a pretraining period where they were gradually exposed to a CCW force field (Figure 4.3D, left). Subjects learned to compensate for the perturbation during these 300 trials. Then, we presented them with both perturbations unpredictably, and again assayed behavior using their errors, actions, success rate, and aftereffects. In the unpredictable period, subjects experienced substantial errors of similar magnitude to both perturbations (Figure 4.3D, right, last 50 trials of unpredictable period, 16.3±2.0° to CCW field, -17.6±1.3° to CW field). They did not maintain the compensation for the CCW field, though they were somewhat biased (Figure 4.3D, left, last 50 trials of unpredictable period, adaptation index, 0.21±0.08, t-test vs. zero, $t(7) = 2.8$, $p = 0.03$) and unsuccessful (last 50 trials of unpredictable period, success rate in non-channel trials, 0.3±0.3%). They also had no aftereffects, assayed in clamp trials (first 25 trials of aftereffects, adaptation index 0.05±0.04, t-test vs. zero, $t(7) = 1.1$, $p = 0.3$) or catch trials (-0.1±1.3°, t-test vs. zero, $t(7) = -0.1$, $p = 0.9$).

Similarly, subjects in Group 1.3 completed a pretraining period where they were gradually exposed to a CW force field (Figure 4.3E, left). After learning to compensate for one perturbation, they were presented with both perturbations in the unpredictable period. They also experienced significant

errors, (Figure 4.3E, right, last 50 trials of unpredictable period, 14.0±1.9° to CCW field, -16.3±0.9° to CW field), did not maintain the compensation for the CW field (Figure 4.3D, left, last 50 trials of unpredictable period, adaptation index, 0.08±0.08, t-test vs. zero, $t(7) = 1.0$, $p = 0.3$), were unsuccessful (last 50 trials of unpredictable period, success rate in non-channel trials, 0.4±0.4%), and had no aftereffects in clamp (first 25 trials of aftereffects, adaptation index -0.06±0.06, t-test vs. zero, $t(7) = -1.1$, $p = 0.3$) or catch trials (-0.05±1.2°, t-test vs. zero, $t(7) = -0.04$, $p = 0.9$).

Overall, the performance of the subjects in Group 1.2 and Group 1.3 demonstrate that even when subjects learned to compensate for one of the perturbations and the motor system was exposed to an action that would produce reward, subjects produced an action consistent with minimization of error in the unpredictable period, even though this did not result in success on the task. In the next experiment, we further tested the hypothesis that exposure to an action that would produce reward may result in maintenance of that action later in the task.

Experiment 2: Do subjects need to reinforce one solution?

Our data in Experiment 1 demonstrated that simply learning to compensate for one of the perturbations was not sufficient to result in persistent selection of a rewarding behavior when both perturbations were presented.

Subjects continued to produce a behavior consistent with minimization of error, even with pretraining. Prior studies have demonstrated that repetition and reinforcement of an action can bias later behavior (Diedrichsen et al., 2010; Huang et al., 2011; Shmuelof et al., 2012). We next hypothesized that subjects might benefit from additional repetition and reinforcement of the successful action.

Subjects in Group 2.1 completed a 5-day experiment (Figure 4.4A). On the first day, they completed the same paradigm as Group 1.2, with pretraining to a CCW field, an unpredictable period where both perturbations were presented, and an aftereffect block. Then, on each of the second to fourth days, they completed 500 trials of additional training with a CCW field. Finally, on the fifth day, they repeated the same paradigm as day 1. We wondered if subjects would be more successful on day 5, as compared to day 1.

On the first day, participants completed the same training as Group 1.2. These subjects did exhibit a bias in their behavior at the end of the unpredictable period (Figure 4.4C, left, last 50 trials of unpredictable period, adaptation index, 0.33±0.05, t-test vs. zero, $t(6) = 6.1$, $p = 9E\text{-}4$), but critically were unsuccessful (last 50 trials of unpredictable period, success rate in non-channel trials, 0.8±0.8%), and had small aftereffects in clamp (first 25 trials of aftereffects, adaptation index 0.14±0.05, t-test vs. zero, $t(6) = 2.8$, $p = 0.03$) and catch trials (-3.0±1.3°, t-test vs. zero, $t(6) = -2.1$, $p = 0.08$). By the fifth day, participants did exhibit an increased

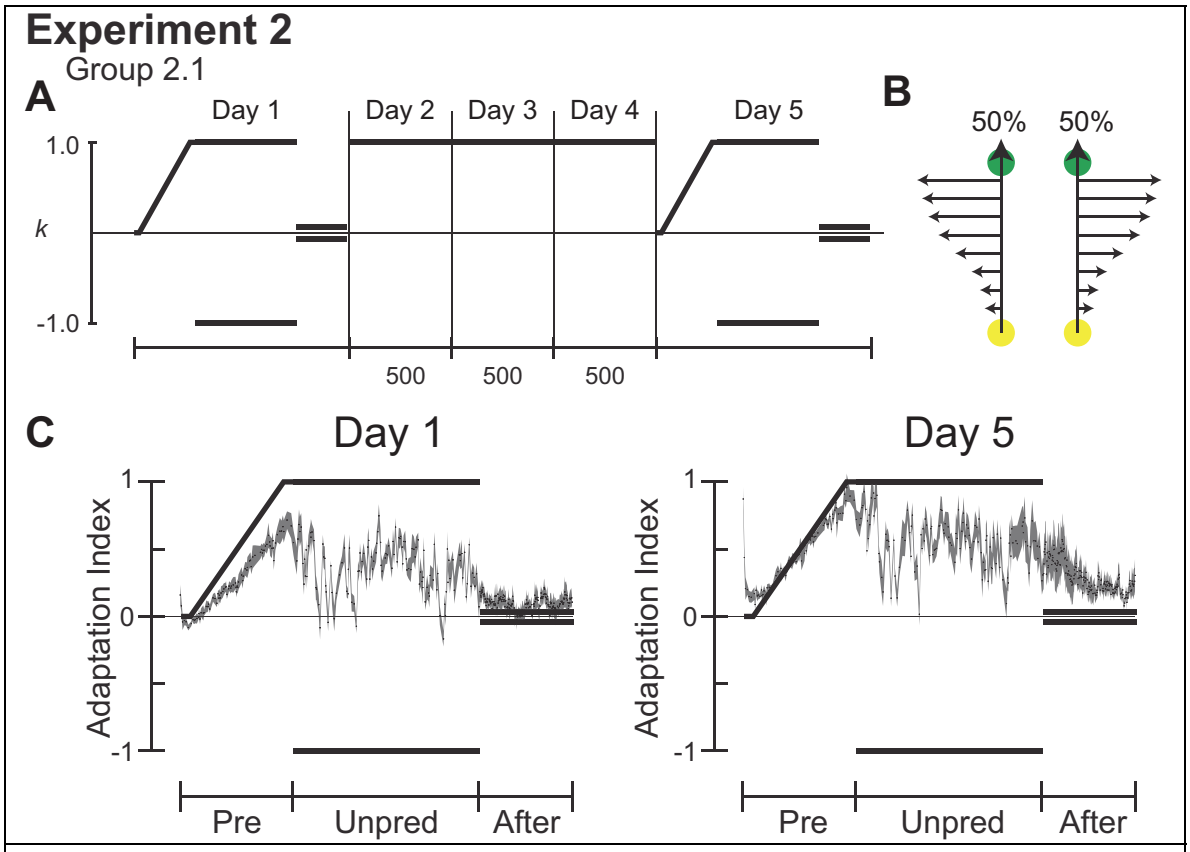

Figure 4.4: Experiment 2: Additional reinforcement. A, In Experiment 2, on days 2, 3, and 4, subjects completed additional trials where they were exposed to a CCW perturbation. B, Leftward and rightward viscous curl fields were presented during the unpredictable period. C, While subjects exhibited a bias in their behavior on Day 5 (right) as compared to Day 1 (left), they did not compensate for one perturbation enough to be successful, and exhibited small aftereffects. Data are mean±SEM across subjects.

bias in their behavior, compared to their behavior on the first day (Figure 4.4C, right, last 50 trials of unpredictable period, adaptation index, 0.55±0.12, paired t-test vs. day 1, $t(6) = 2.6$, $p = 0.04$), with somewhat larger aftereffects (first 25 trials of aftereffects, adaptation index in clamp trials: 0.42±0.13, paired t-test vs. day 1, $t(6) = 2.1$, $p = 0.08$; aftereffects in catch trials: -6.7±1.5°, paired t-test vs. day 1, $t(6) = -2.05$, $p = 0.09$). However, the bias in the behavior was not accompanied by an increased rate of reward. Subjects were no more successful on the fifth day than they were on the first (last 50 trials of unpredictable period, success rate in

non-channel trials, 5±2%, paired t-test vs. day 1, $t(6) = -2.1$, $p = 0.09$). Repetition and reinforcement of an action biased behavior, but did not result in additional success. This bias may be due to repetition of the action (Diedrichsen et al., 2010).

Experiment 3: Can we make the error signal less salient?

In Experiment 1 and Experiment 2, we observed that subjects produced actions that minimized the cost of error, even when this impaired their ability to succeed at the task. Next, we wondered if subjects might choose to be successful if the error signal were less salient. To do so, we modulated the feedback subjects received in two ways. Subjects experience errors from two sensors in this task: visual error, due to the cursor missing the target, and proprioceptive error, due to the hand being perturbed from its trajectory to the target. In Group 3.1, we attempted to reduce the visual error; in Group 3.2, we eliminated proprioceptive errors.

In Group 3.1, we reduced the visual error experienced by subjects by hiding the visual cursor during part of the movement, and only providing subjects with endpoint visual feedback (Figure 4.5A, left). Subjects again completed pretraining (CCW field), unpredictable, and aftereffect blocks. Even with a reduction in visual feedback, subjects exhibited large error to both perturbations (Figure 4.5A, right, last 50 trials of unpredictable period, 17.0±2.2°

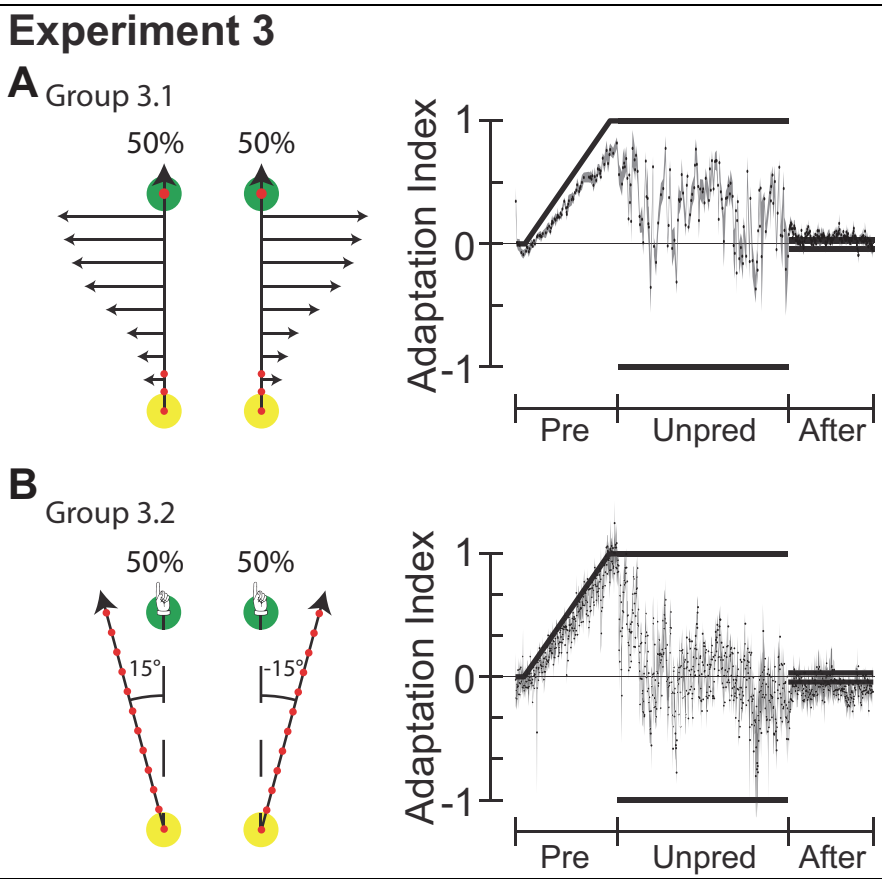

Figure 4.5: Experiment 3: reducing error. In Experiment 3, the visual and proprioceptive feedback was manipulated. A, In Group 3.1, midmovement feedback was eliminated (left, red indicates cursor). With limited visual feedback, subjects did not maintain their behavior, did not achieve success in the unpredictable period, and did not show aftereffects (right). B, In Group 3.2, proprioceptive errors were eliminated by using visual perturbations. The cursor was visible throughout the movement. Subjects were pretrained to a CCW perturbation. Even without proprioceptive errors, subjects produced a behavior consistent with minimization of error in the unpredictable period. They were unsuccessful and displayed no aftereffects. Data are mean±SEM across subjects.

to CCW field, -18.2±1.7° to CW field) and did not completely compensate for one of the perturbations, though they exhibited some bias, (last 50 trials of unpredictable period, adaptation index, 0.26±0.06, t-test vs. zero, $t(7) = 4.1$, $p = 4E\text{-}3$). They were unsuccessful (last 50 trials of unpredictable period, success rate in non-channel trials, 2±1%), and exhibited small aftereffects in clamp (first

25 trials of aftereffects, adaptation index 0.10±0.04, t-test vs. zero, $t(7) = 2.6$, $p = 0.03$) and catch trials (1.3±1.2°, t-test vs. zero, $t(7) = 1.0$, $p = 0.4$).

In Group 3.2, we eliminated the proprioceptive errors in the task by using visual perturbations. In this paradigm, no force was applied to the hand, and only a visuomotor rotation was used to perturb the cursor. The position of the cursor was rotated +15° and -15° in the unpredictable period (Figure 4.5B, left). These rotations were similar in magnitude to the visual errors experienced by the force field groups (e.g. Group 1.2, CW field: -17.6±1.3°; CCW field: 16.3±2.0°). In this paradigm, the hand is not perturbed, and so no proprioceptive errors should be experienced. Subjects in this group learned to compensate for a CCW rotation in the pretraining period (Figure 4.5B, right). However, when presented with both perturbations unpredictably, they did not maintain that compensation (last 50 trials of unpredictable period, adaptation index -0.17±0.09, t-test vs. zero, $t(7) = -1.9$, $p = 0.10$) and were unsuccessful (last 50 trials of unpredictable, success rate in non-channel trials, 1.4±1.0%). They also exhibited no significant aftereffects (first 25 trials of aftereffects, adaptation index -0.15±0.08, t-test vs. zero, $t(7) = -1.91$, $p = 0.10$).

Subjects selected an action that was consistent with minimization of error, and inconsistent with achieving success on the task, even when visual error was reduced and proprioceptive errors were eliminated. Learning from error appeared to continue to dominate behavior in this task.

Experiment 4: Does reward matter in the task?

Our results in Experiments 1 to 3 demonstrated a strong role for error over reward in motor learning, assessed by presentation of two perturbations unpredictably. These results beg the question: does success on the task matter to subjects? Subjects received a target 'explosion' and a point added to their score if the cursor intersected the target in the appropriate amount of time. Perhaps subjects did not achieve task success because the rewards were not meaningful.

To test whether reward mattered, we again provided subjects with two perturbations, but attempted to do so without creating persistent sensory prediction errors (SPEs). In Group 4.1, subjects pretrained to a CCW visuomotor rotation with a single cursor. Then, in the "unpredictable" period, we presented two cursors on each trial (Figure 4.6A, left). One cursor was perturbed +15° (CCW), and the other cursor was perturbed -15° (CW). Subjects received feedback indicating success if either cursor hit the target. Critically, in this experiment both cursors, and hence both perturbations, were presented simultaneously.

What behaviors would be expected in this experiment, when an action produces two consequences simultaneously? In this task, we can think about error in two ways – as prediction error and as performance error. Prediction errors are the difference between the expected and observed feedback. Performance errors are the difference between the desired and observed

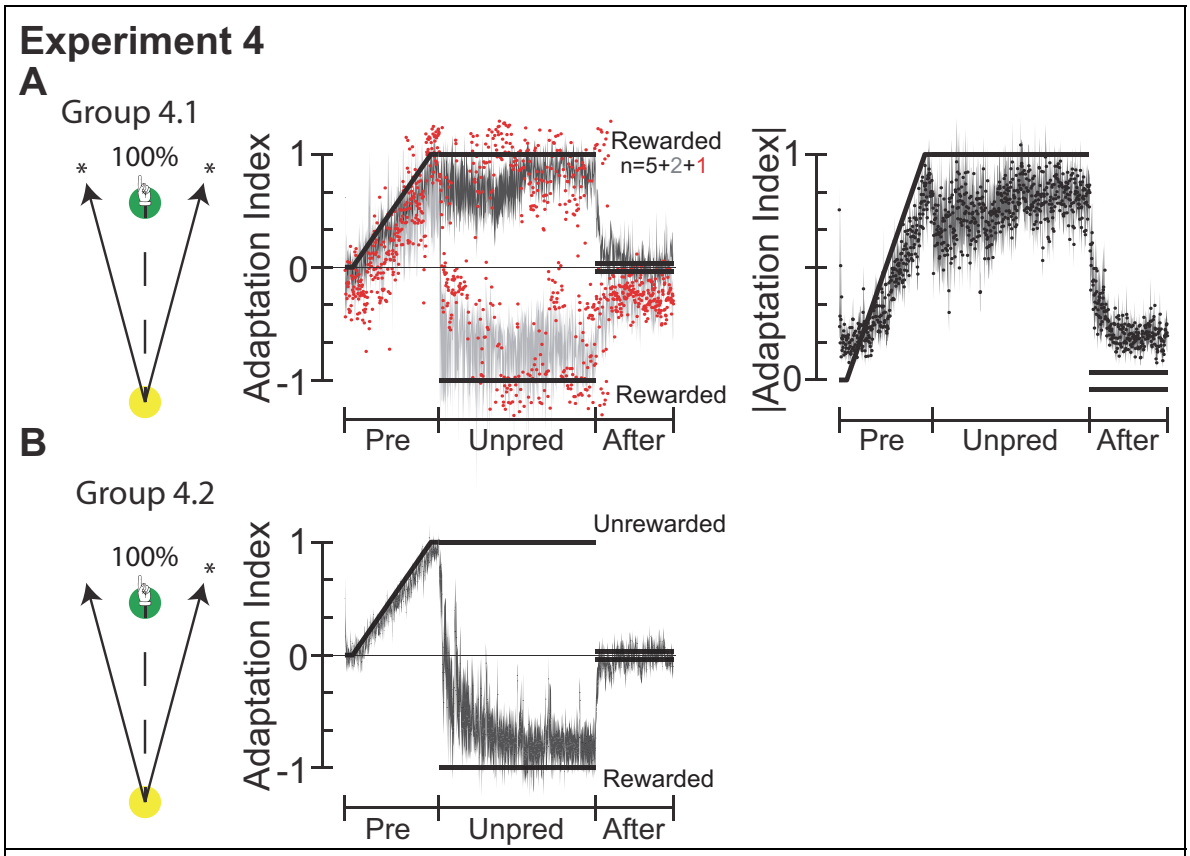

Figure 4.6: Experiment 4: The role of reward in the absence of sensory prediction error. A, In Group 4.1, subjects were presented with both visual perturbations simultaneously (left). If either cursor hit the target, subjects received feedback indicating success (denoted by *). Of the 8 subjects in this group, 5 maintained the pretrained behavior that compensates for a CCW rotation; 2 produced behaviors that compensated for a CW rotation, and 1 alternated between these behaviors (middle). The absolute adaptation index (right) captures these behaviors together. Subjects maintained a successful action during the "unpredictable" period and exhibited significant aftereffects. B, In Group 4.2, we pretrained subjects to compensate for a CCW rotation, and then only rewarded the CW perturbed cursor in the unpredictable period (denoted by *). Subjects produced the rewarding action and were successful, in the absence of persistent sensory prediction errors. Data are mean±SEM across subjects.

feedback. Because we assumed that subjects desired, in terms of error, to place the cursor in the center of the target and selected an action such that the expected consequences will be the desired ones, prediction and performance errors could not be disambiguated in prior experiments. We, hence, referred to them collectively as "error". In this experiment, in contrast, these quantities can be dissociated in some cases.

Suppose subjects sought to minimize the performance error of both cursors in the task. Then, we would expect to compensate for the mean perturbation, move directly to the target, and be unsuccessful. This results in small performance errors of equal and opposite signs from each of the two perturbations (Figure 4.1C).

Alternatively, suppose subjects used prediction errors to learn the perturbations applied to both cursors. In this experiment, a single perturbation is applied consistently to each cursor. If subjects are able to represent the perturbations to both cursors simultaneously in the brain, then, by the end of the unpredictable period, they would experience no prediction error. In this case, if reward matters in the task, we would expect subjects to compensate for one of the two perturbations, placing one cursor in the target.

In Figure 4.6A (middle), the behavior of subjects is shown. Subjects adopt one of three behaviors: some subjects (5 of 8) maintained their compensation for the CCW perturbation; some subjects (2 of 8) compensated for the other CW perturbation; and one subject sometimes compensated for the CCW and sometimes compensated for the CW perturbation. However, in all cases subjects appeared to compensate for the perturbations, and select a rewarding action. To group these subjects, we plotted the absolute value of the adaptation index, so that an adaptation index of -1 (which indicates an action which was successful) is grouped with an adaptation index of +1 (which also indicates an action which

was successful). We see that subjects exhibited a large bias in their behavior (Figure 4.6A, right, last 50 trials of unpredictable period, absolute adaptation index, 0.80±0.05, t-test vs. zero, $t(7) = 16.9$, $p = 6E\text{-}7$), were successful in the unpredictable period (last 50 trials of unpredictable period, success rate in non-channel trials, 32±8%), and exhibited significant aftereffects (first 25 trials of aftereffects, absolute adaptation index 0.48±0.06, t-test vs. zero, $t(7) = 8.4$, $p = 7E\text{-}5$). These results suggested that subjects were able to predict both consequences of their actions, and that prediction error, not performance error, was involved in this learning task. This use of prediction error, as opposed to performance error, is consistent with prior work (Mazzoni and Krakauer, 2006; Taylor and Ivry, 2011). Furthermore, in the absence of persistent prediction errors, even when presented with two perturbations, subjects were able to select or maintain a rewarding action.

We wanted to further test if subjects were attempting to be successful in the task. Perhaps subjects in Group 4.1 simply sought to place a cursor in the target and task success did not play a role in the selection of their behavior. These goals are similar, but in Group 4.1, because placing either cursor in the target always produced feedback indicating success, we could not disambiguate a desire to hit the target from a desire to be explicitly successful. To further test if task success mattered to participants in this task, we recruited an additional group of subjects.

Participants in Group 4.2 were also pretrained with a CCW visuomotor rotation. Then, in the unpredictable period, they were presented with two sensory consequences for each action – one cursor was perturbed 15° CCW and one cursor was perturbed 15° CW. However, in this case, we only provided success feedback if the CW perturbed cursor hit the target (Figure 4.6B, left). Critically, at the end of the pretraining and beginning of the unpredictable period, because of the CCW pretraining, subjects were placing the CCW perturbed cursor in the target. If participants sought only to place a cursor in the target, and task success did not play a role in the selection of action, we would expect some subjects to maintain the behavior from the pretraining, compensating for the CCW perturbation (as in Group 4.1). On the other hand, if subjects in Group 4.1 and this group sought to achieve task success, then these subjects should all switch their behavior and compensate for the CW perturbation. Indeed, the latter is what we observed (Figure 4.6B, right). By the end of the unpredictable period, subjects compensated for the CW perturbation (last 50 trials, adaptation index, -0.80±0.14, t-test vs. zero, $t(7) = -5.6$, $p = 8E\text{-}4$), were successful (last 50 trials of unpredictable period, success rate in non-channel trials, 38±7%), and exhibited aftereffects that trended towards significance (first 25 trials of aftereffects, adaptation index -0.22±0.10, t-test vs. zero, $t(7) = -2.3$, $p = 0.055$).

In both groups in this experiment, we eliminated the persistent sensory prediction errors in the "unpredictable" period, by provided the two perturbations simultaneously. In doing so, we observed that subjects did produce an action consistent with the maximization of task success, in the absence of persistent sensory prediction errors. In the next experiment, we attempted to use this paradigm to further modulate sensory and reward prediction errors in this task.

Experiment 5: Can we modulate sensory and reward prediction errors?

In Experiment 5, we presented both perturbations simultaneously, as in Experiment 4, but sought to modulate the sensory and reward prediction errors in the task. In Group 5.1, we used color to introduce a mild sensory prediction error (SPE). As in prior experiments, subjects were pretrained to a CCW visuomotor rotation. Then, in the unpredictable period, they were presented with two cursors, so that both perturbations were presented simultaneously. As in Group 4.1, we rewarded actions that would move either cursor to the target. However, we added a mild SPE in the form of a coloring of the cursor. One cursor was white, as before, and the other was colored gray (Figure 4.7A, left). The coloring was applied pseudorandomly, so that subjects experienced a mild SPE. The color of a given cursor (or the location of the cursor of a given color) was unpredictable. However, success feedback was given if either cursor hit the target, so there was no reward prediction error (RPE) – any action that resulted

in success would always result in success. In this paradigm, subjects would experience a mild, persistent SPE, but no RPE.

Subjects in Group 5.1 did compensate for the perturbations (Figure 4.7A, right, last 50 trials of unpredictable period, absolute adaptation index, 0.69±0.09, t-test vs. zero, $t(7) = 8.0$, $p = 9E\text{-}5$), were successful in the unpredictable period (last 50 trials of unpredictable period, success rate in non-channel trials, 32±9%), and exhibited significant aftereffects (first 25 trials of aftereffects, absolute adaptation index 0.44±0.06, t-test vs. zero, $t(7) = 7.6$, $p = 1E\text{-}4$). That is, with only a mild, persistent sensory prediction error, subjects did select an action that resulted in success.

In prior Experiments 1 to 3, subjects experienced both persistent SPEs and RPEs in the unpredictable period. In those experiments, an action would result in disparate sensory consequences, but would also only result in reward at most 50% of the time. As a result, if a successful action were made, subjects would experience reward prediction errors – a lack of success when success would be expected. In contrast, in Experiments 4 and 5.1, because a single action could produce reward 100% of the time, there was no RPE. We next wondered how subjects would behave in the presence of the mild SPE with a similar RPE. To do so, we again used coloring of the cursor to introduce a mild SPE for subjects in Group 5.2. Subjects completed a paradigm similar to that completed by Group 5.1, except we added an additional RPE by providing success feedback only if the

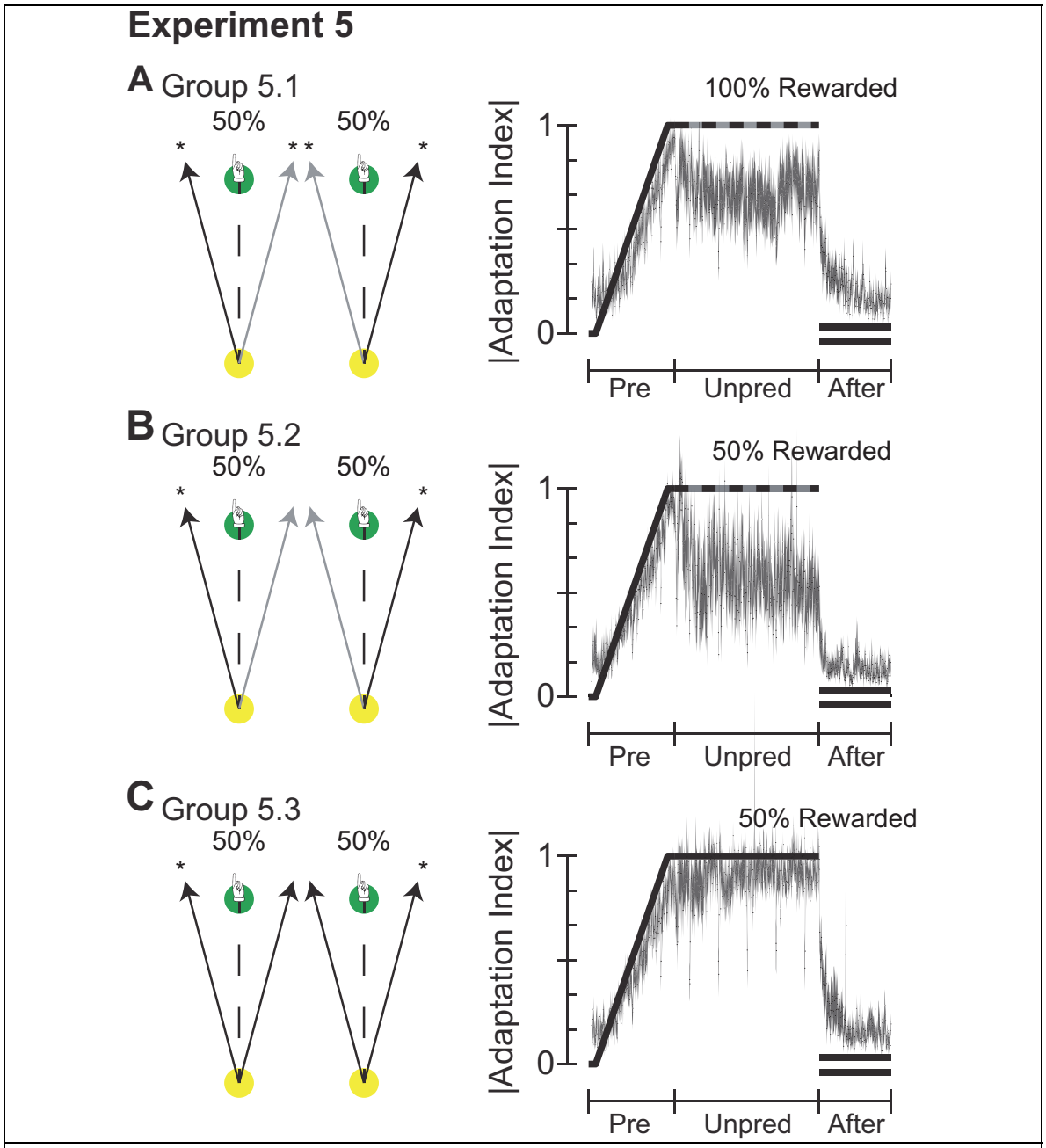

Figure 4.7: Experiment 5: Sensory and reward prediction error. In Experiment 5, we manipulated the sensory and reward prediction errors in the task. A, In Group 5.1, we presented subjects with a mild sensory prediction error in the unpredictable period, by randomly coloring one cursor (left). If either cursor hit the target, subjects received feedback indicating success (denoted by *). In this case, subjects maintain the rewarding action, and exhibit aftereffects (right). B, In Group 5.2, we presented subjects with the same mild sensory prediction error, but also a reward prediction error by only rewarding the white cursor (left). In this case, subjects exhibited biased behavior, but were not successful. C, In Group 5.3, we presented subjects with a reward prediction error, but no sensory prediction error. Both cursors were white, and only one cursor was rewarded, pseudorandomly. In this case, subjects maintain the rewarding behavior, were successful, and exhibited significant aftereffects. Data are mean±SEM across subjects.

white cursor hit the target (Figure 4.7B, left). In this case, subjects' behavior was biased (Figure 4.7B, right, last 50 trials of unpredictable period, absolute adaptation index, 0.48±0.09, t-test vs. zero, $t(7) = 5.3$, $p = 1E\text{-}3$), but unsuccessful on the whole (last 50 trials of unpredictable period, success rate in non-channel trials, 7±3%). That is, in the presence of both sensory and reward prediction errors, subjects did not overcome the minimization of error to select a successful action.

Finally, we wondered if reward prediction error alone in this and other experiments was sufficient to result in behavior that was unsuccessful. So, in Group 5.3, we presented subjects with an RPE, but no SPE. Subjects were pretrained to a CCW rotation. Then, in the unpredictable period, they received two cursors, both of which were white. As in Group 4.1 and 4.2, there was no persistent SPE. However, subjects were successful only if one of the two cursors, selected pseudorandomly, hit the target (Figure 4.7C, left). In this case, a single action could only result in reward 50% of the time, and so subjects would experience a persistent RPE if they selected a successful action. Despite this, subjects did select a rewarding action – their behavior compensated for the perturbation (Figure 4.7C, right, last 50 trials of unpredictable period, absolute adaptation index, 0.90±0.02, t-test vs. zero, $t(7) = 41.3$, $p = 1E\text{-}9$), they were successful (last 50 trials of unpredictable period, success rate in non-channel trials, maximum 50%, 25±1%), and they exhibited significant aftereffects (first 25

trials of aftereffects, absolute adaptation index 0.49±0.07, t-test vs. zero, t(7) = 7.4, p = 2E-4). Subjects selected an action that maximized task success when a persistent RPE is present, in the absence of SPEs.

We conducted an ANOVA to directly test for the effect of persistent sensory and reward prediction errors on the rate of success in the unpredictable period of the task. We used the four groups where two cursors were presented and the presence of SPEs and RPEs was manipulated: Group 4.1 (SPE absent, RPE absent), Group 5.1 (SPE present, RPE absent), Group 5.2 (SPE present, RPE present), and Group 5.3 (SPE absent, RPE present). There was a significant main effect of the presence of a persistent SPE ($F(1, 1) = 17.6$, $p = 3E\text{-}4$), no significant main effect of RPE ($F(1, 1) = 0$, $p = 0.96$), and a significant SPE*RPE interaction ($F(1, 1) = 11.8$, $p = 2E\text{-}3$). That is, in the absence of sensory prediction errors subjects selected an action that resulted in task success. Furthermore, when SPEs were mild and there was no RPE, subjects could balance the two systems and be successful. But when persistent SPEs were accompanied by RPEs, sensory prediction error dominated motor learning – subjects selected actions which minimized error, forsaking reward.

Summary

Figure 4.8 summarizes the results of all of the experiments. We plotted the rate of success and hit rate in the final 50 trials of each epoch. For

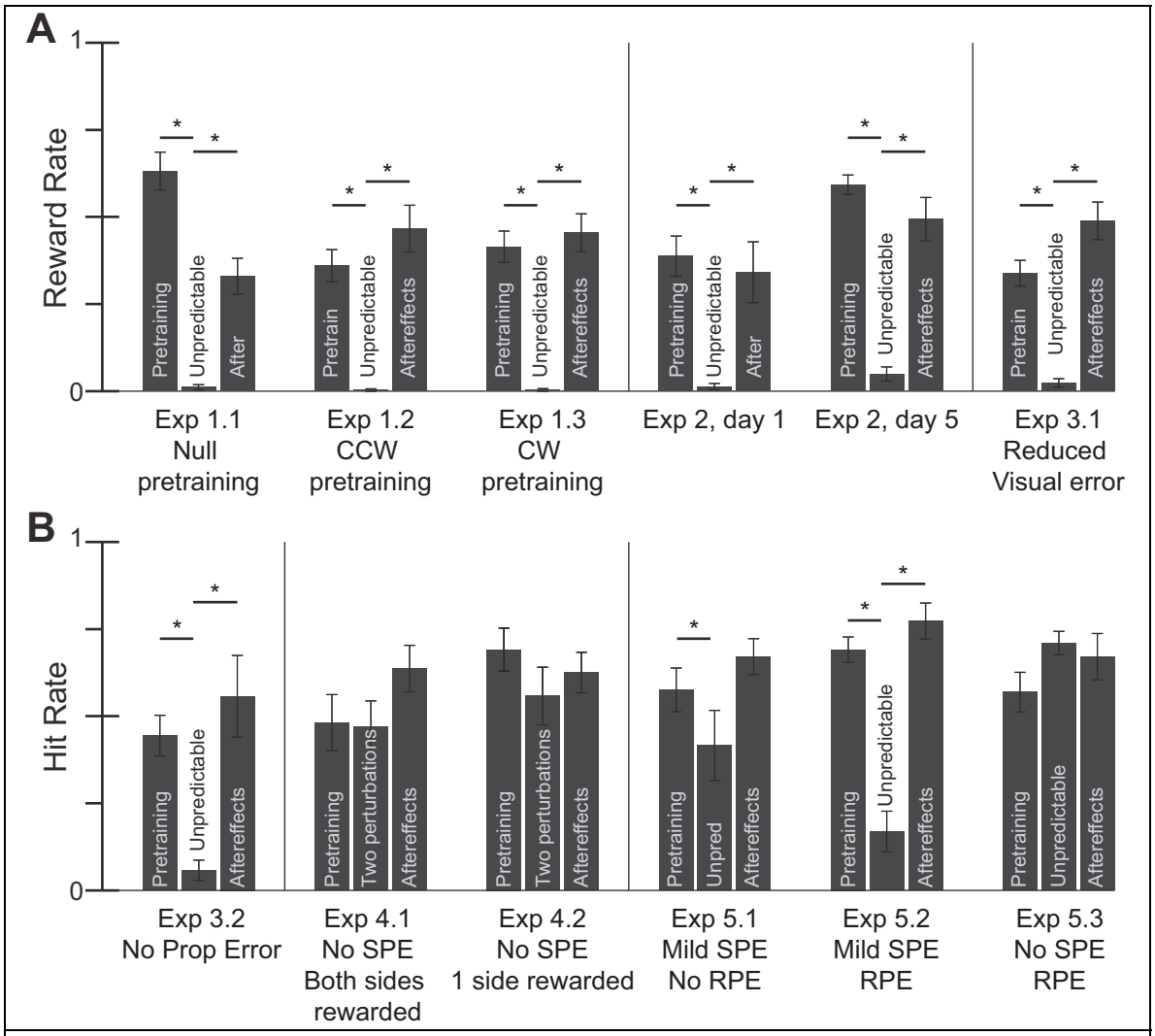

Figure 4.8: Summary of behavior. A, The reward rate in force field experiments, the fraction of perturbation trials that were successful. B, The hit rate in visuomotor rotation experiments, the fraction of perturbation trials where a cursor hit the target. The mean reward and hit rate from the final 50 trials of each epoch are shown. Data are mean±SEM across subjects. Paired t-tests conducted between pretraining and unpredictable; and unpredictable and aftereffects epochs. * indicates $p < 0.05$

experiments where force perturbations were applied, we compared the rate of success at the end of the pretraining, unpredictable, and aftereffect periods (excluding error-clamp trials, Figure 4.8A). For experiments where visual perturbations where applied, we plotted the hit rate, the proportion of trials in which subjects were successful, or would have been success had the other

perturbation been present and rewarded (Figure 4.8B). Overall, we observed that in each experiment where persistent SPEs were absent in the unpredictable period (Exps 4, 5.3), or when SPEs were mild and persistent RPEs were absent (Exp 5.1), subjects produced rewarding actions. However, when persistent SPEs and RPEs were both present (Exps 1-3, 5.2), subjects did not select behaviors that resulted in task success.

Section 4.4: Discussion

When the motor system plans and executes an action, it receives feedback from the body's sensors about the consequences of that action. When actions produce undesirable outcomes – errors or a lack of success – the brain can refine its subsequent behavior in order to reduce error and increase the probability of reward. These two systems of learning are used to develop models of the world around us, adapt to perturbations, and produce accurate movements (Thoroughman and Shadmehr, 2000; Izawa and Shadmehr, 2011). In many cases, they provide congruous information – actions which minimize error also maximize the rate of task success. In some cases, however, the error-based and reward-based systems provide incongruous information about the desirable actions – when actions which minimize error are distinct from those that produce reward. When this is the case, how does the brain and motor system integrate the conflicting information to select an action?

In the work in this chapter, we presented subjects with a paradigm where actions which would result in task success were distinct from those which would minimize the error associated with the sensory consequences of the action. That is, we pitted the error-based and reward-based learning systems against each other. We found that learning from sensory prediction errors dominated learning, even though reward did influence behavior (when persistent SPEs were absent). When the SPE was mild, without RPEs, subjects were able to overcome the error-based learning system and produce a successful behavior; otherwise learning from SPEs appeared to control behavior, even though this resulted in failure on the task.

Our results expand on prior studies demonstrating a strong role for sensory prediction error in the selection of actions. Mazzoni and Krakauer (2006) and Taylor and Ivry (2011) presented subjects with a visuomotor rotation perturbation, but also explicitly presented subjects with a cognitive strategy to compensate for that perturbation. Using the strategy, subjects were immediately successful and did not experience performance errors, but likely experienced implicit sensory prediction errors. The authors observed that subjects did not initially maintain the successful, strategic behavior. Instead, subjects learned from the prediction errors, transiently forsaking reward. After approximately 100 trials of training with both the perturbation and strategy, however, subjects returned to a behavior that produced reward, presumably because they had

updated their internal model of the perturbation using prediction errors. That is, eventually they were producing an action that minimized sensory prediction errors and also resulted in reward. These results demonstrated a dominant role for sensory prediction over task success where these two goals were transiently at odds. In our work, we wondered if subjects would be able to overcome the error-based learning system when these systems were persistently at odds, for hundreds of trials. The results in this chapter demonstrate that even when sensory prediction error and task success were persistently at odds, learning from error dominated behavior. Furthermore, even when subjects were taught a successful behavior implicitly with training (as opposed to explicitly), error-based learning systems still dominated. The return to successful behavior in prior work was likely due to compensation for sensory prediction errors, and not due to the lack of reward.

In Experiments 4 and 5, we presented subjects with two perturbed cursors – two sensory consequences for each action. Kasuga et al. (2013) used similar feedback to study single trial learning. In that work, subjects were presented with two identical cursors on each trial, as in Experiments 4.1, 4.2, and 5.3. The authors presented several combinations of perturbations to the cursors, randomly, and observed how much subjects changed their behavior in response to a single pair of perturbations. In particular, in some trials they presented one unperturbed cursor, and another cursor with a perturbation of ±30°. This

condition produces feedback similar to the feedback subjects received in our work early in the unpredictable period – one cursor moved to the target, and the other moved 30° away. Kasuga et al. (2013) found that subjects did adapt their behavior on the next trial in this case. In contrast, we found that when exposed to this feedback persistently, subjects maintained their behavior. However, sensory prediction errors are eliminated only after repeated exposure to this feedback; we observed maintenance of behavior in the absence of sensory prediction errors, while Kasuga et al. observed changing behavior in the presence of sensory prediction errors, for the same type of feedback. Taken together, these results support our conclusion that sensory prediction errors drive the selection of behavior, even when two sensory consequences are observed.

It is important to note that, while we observe a dominant role for error-based learning, our results do not preclude reward-based learning in motor control. In fact, the results of Experiment 4 demonstrate a compelling role for reward-based learning, when prediction errors are not available to the error-based system. Learning from sensory-prediction error may be indifferent to the presentation of reward, as we observed in Chapter 3, but in the absence of SPEs, subjects did value task success. In general, though, learning from sensory prediction error dominated learning and behavior.

Chapter 5: Vigor of movements and the temporal discounting of reward[3]

Section 5.1: Motivation

Among healthy people, there are similarities in how we walk, reach, or move our eyes. To explain these regularities, theories have suggested that the nervous system produces motor commands to minimize metabolic costs (Hoyt and Taylor, 1981; Willis et al., 2005) or kinematic variability (Harris and Wolpert, 1998). Yet, these theories cannot explain the fact that people (Xu-Wilson et al., 2009) and other primates (Kawagoe et al., 1998; Takikawa et al., 2002; Opris et al., 2011) move sooner or faster when there is an opportunity to acquire a greater amount of reward. For example, people produce saccades that have higher velocities in environments that offer greater rate of reward (Haith et al., 2012). In addition, people with Parkinson's disease, who exhibit differences in the way they value reward (Frank et al., 2004) also make slower movements than healthy people of the same age (Mazzoni et al., 2007). These observations suggest that in addition to efficiency and variability, the reward landscape and the way rewards are valued affects the speed with which we move (Niv et al., 2007).

[3] The work in this chapter has been published in the Journal of Neuroscience: Choi JES*, Vaswani PA*, Shadmehr R (2014) Vigor of movements and the cost of time in decision making. J Neurosci 34:1212–1223.

In principle, why should reward affect speed of movements? If we assume that the purpose of any movement is to arrive at a more rewarding state, then movement duration carries a cost because passage of time discounts reward. That is, it is better to receive the reward sooner rather than later. Therefore, a movement that takes longer to complete produces a greater devaluation of reward. Motor commands that guide a movement may be a balance between a desire to reduce inaccuracy (move slowly and improve precision), and a desire to maximize reward (move quickly and get reward sooner) (Shadmehr et al., 2010).

Suppose that we have two subjects who are similar in terms of biomechanics of their body, but who temporally discount reward differently. The theory predicts that the subject who discounts reward steeply should generally move faster (Shadmehr and Mussa-Ivaldi, 2012). Indeed, in populations in which development or disease affects temporal discounting there are between-population differences in saccade velocity (Shadmehr et al., 2010). The critical question, however, is whether the between-subject differences in movement are related to between-subject differences in discounting of reward.

A temporal discount function can be measured in scenarios in which subjects compare a rewarding state that can be attained soon, with a more rewarding state that can be attained later (Millar and Navarick, 1984; Myerson and Green, 1995). Here, we measured saccadic eye movements and observed

that some people moved their eyes with a peak velocity that was 50% faster than others. This difference was consistent in repeated measurements, appearing to be a trait. We then estimated temporal discounting in a decision-making task in which people decided how long to wait in order to improve their odds of success. In our task, every choice resulted in a real and immediate consequence, reinforcing the choice and affecting subsequent choices. We found that people who made faster movements, as evidenced by saccade velocities, also tended to have a steep temporal discount function, as evidenced by the shorter periods of time they chose to wait to obtain additional reward.

Section 5.2: Methods

Subjects sat in a darkened room in front of a CRT monitor (36.5 x 27.5 cm, 1024 x 768 pixel, light grey background, frame rate 120Hz) with head restrained using a dental bite bar. Visual targets (black, diameter = 1 deg) were presented on a CRT monitor with Matlab 7.4 (Mathworks) using Psychophysics Toolbox 3. The screen was placed at a distance of 31 cm from the subject's face and an Eyelink 1000 (SR Research) infrared camera recording system (sampling rate = 1000 Hz) was used to record movement of the right eye. The experiments were approved by the Johns Hopkins Institutional Review Board. Volunteers were healthy with no known neurological disorders.

We wished to answer two questions: 1) how much did movement vigor, as measured by peak saccade velocity as a function of amplitude, vary across

healthy individuals; and 2) was an individual's temporal discounting of reward as measured in a decision making task a predictor of that individual's movement vigor. N = 23 volunteers (14 females, 26.9±6.8 years old, mean±SD) participated in our two-part study, which were conducted on two separate days.

Movement vigor

In this part of the experiment we wished to determine the range of movement speeds across our population of healthy individuals. We measured kinematics of saccadic eye movements, determined the within subject reproducibility of these movements, and the between subject differences. Targets that were 5°, 10°, 15°, 20°, 25°, 30°, 35°, or 40° apart on the horizontal axis were presented on a CRT monitor, centered on the midline of the right eye. Target amplitudes were ordered pseudo-randomly in a block-wise fashion. Each target was presented 30 times in a row; resulting in 29 saccades (we discarded the first saccade as this saccade was from a midline location to the first target and therefore was half the target amplitude).

A trial began with display of a fixation spot. Our instructions were: "A sequence of targets will appear on the screen. Please look at each target and maintain fixation until you see the next target." Each target was displayed for 1 sec plus a random time distributed uniformly over -100 to 100 ms (Figure 5.1A). Appearance of the target acted as a go-cue. We did not enforce any gaze

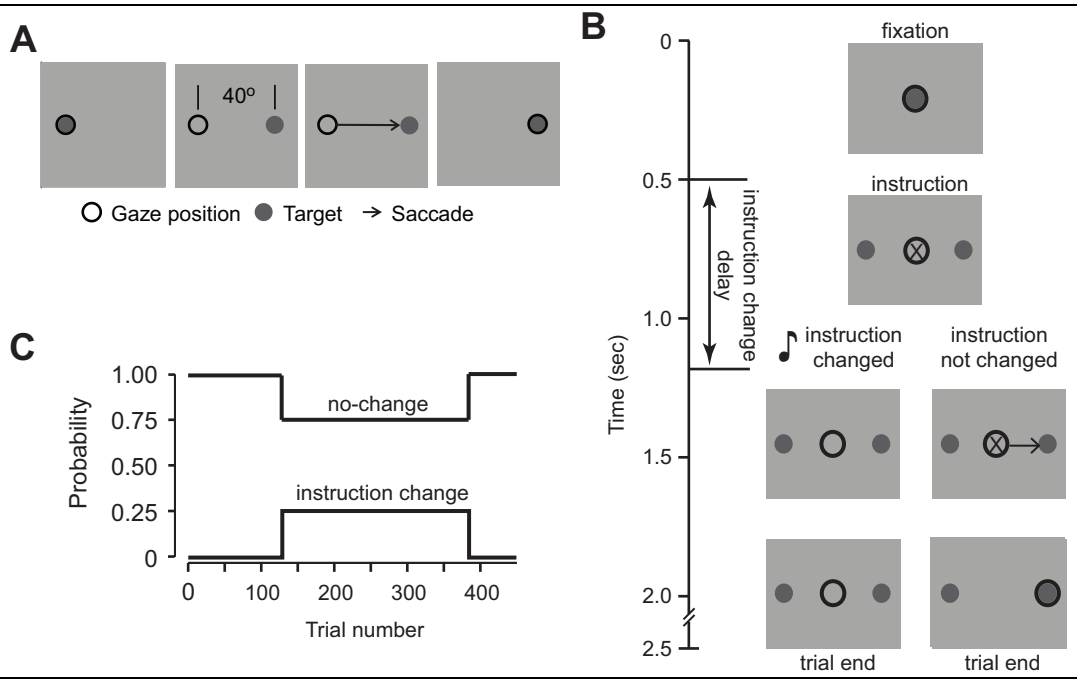

Figure 5.1: Experimental protocol. **A.** Part 1 of the experiment: measurement of saccade vigor. The trial began with a fixation spot of 0.5°, and then presentation of a target spot of 0.5° at a given displacement along the horizontal axis. Targets were presented for 1 second plus or minus a random time period. The targets were centered about the midline of the right eye. **B.** Part 2 of the experiment: measurement of temporal discount function. The trial began with a central fixation spot. Two targets were presented at 20° from fixation along with an instruction at the fixation spot indicating which target was the direction of the correct saccade. In blocks 3-6s, there was a 25% probability that following a variable delay period a second instruction would be given, indicating the previously instructed saccade should be canceled. The delay period was adaptively adjusted to the success and failure of the subject on previous trials: success made the delay period 30 ms longer. The experiment attempted to measure the length of time the subject was willing to wait to improve their probability of success. **C.** Schedule of instruction probabilities in part 2 of the experiment.

precision requirements; subjects received only the verbal instruction to look at the targets. The subjects received a short break after completion of two target amplitude blocks.

To assess the reproducibility of our results, five of the subjects were examined repeatedly on this task on 4 separate days. To ensure that time of day was not a factor, we selected four test times during the day (ranging from early

morning to late afternoon) and tested each of these 5 subjects once on each test time.

To define vigor, we considered the peak velocity of saccades as a function of saccade amplitude. We measured amplitude via endpoint displacement of the eye, with positive displacement indicating temporal saccades, and negative displacement indicating nasal saccades. Saccade peak velocity tends to increase as amplitude increases and saturates around 30°. As we will show, the between subject differences in the velocity-amplitude function is accurately summarized by a scaling factor. Let us label the across-subject mean of the velocity-amplitude function as $g(x)$, where x is endpoint displacement and $g(x)$ describes the relationship between displacement and average velocity across the population. That is,

$$g(x) = E\left[v(x)\right] \qquad \text{Eq. 5.1}$$

In Eq. 5.1, $E[\ ]$ is the expected value operator, computing the across subject mean of the velocity-amplitude relationship. We will show that each subject's velocity-amplitude relationship is a scaled version of this function. That is, for subject i, peak velocity at displacement x is described by:

$$v_i(x) = \alpha_i g(x) \qquad \text{Eq. 5.2}$$

The scale factor α_i is our proxy for vigor of saccades for subject i.

Saccade beginning and end were marked using a 30 deg / sec velocity threshold (held for at least 4 ms). We used the following criteria to accept a saccade: no blinking during the saccade, displacement less than 100 deg, peak velocity less than 1500 deg/s.

Temporal discount function

There are two classes of experiments that are used to measure temporal discounting in humans (Navarick, 2004). In one class, subjects are presented with potentially rewarding outcomes, the resulting choice is measured, and the consequences are immediately applied. The key element of this "operant class" of experiments is that the choices have real and immediate consequences that are experienced before any other choices are made. These consequences act as reinforcements or punishments, which then affect the next choice. All experiments in animals and some experiments in humans (Jimura et al., 2009) are of this class.

In the "non-operant" class of experiments, rewarding states are presented (often small amounts of money soon vs. larger amounts later), a choice is measured, but the consequences of that choice are not experienced before the next choice is made. This is because the delay associated with the two rewarding states is typically days or weeks (rather than seconds, as in the operant experiments). Furthermore, nearly all rewards are hypothetical. However,

participants are sometimes instructed that a couple of their choices will be selected for real payment after the session. Importantly, because all decisions are made before the money is received; the reward or punishment is not a reinforcement that affects the subsequent choices that are made in the experiment.

Here, we designed an experiment to measure temporal discounting that relied on the operant procedure. Every choice produced immediate and real consequences that in case of success could positively reinforce the choice, and in case of failure could negatively reinforce the choice. Using a model (described later), we predicted how the consequence of each choice would affect the subsequent choice, and how this trial-to-trial effect would be a proxy for the steepness of the temporal discount function.

Let us explain our task first intuitively, and then in a mathematical framework (in Section 5.3). Imagine joining a line where one has to wait to experience an event. We join the line with a prior belief regarding how long we have to wait. In our task, we control this prior belief by manipulating the history of when that event takes place. As we wait in line, with passage of time we update our expectation of how much longer we have to wait. At some point, we may decide that the waiting is not worthwhile and leave the line. According to our model, the time at which we abandon the line is the time at which the temporally discounted value of reward has reached and passed a local

maximum. The time when we abandon the line is a measure that will act as a proxy for the steepness of the temporal discount function.

Our task is shown in Figure 5.1B. Subjects were instructed to look at the central fixation spot (0.5°) presented for 500 ms. Subjects were instructed: "If the central fixation spot turns into an X, move your eyes to look at the target on the right. If the central fixation point turns into an O, move your eyes to look at the target on the left." Next, we presented two visual targets of size 0.5° at ±20°, along with an instruction at the fixation spot indicating which target the subject should saccade to: an "X" instructed a saccade to the right target, and an "O" instructed a saccade to the left target

The experiment consisted of seven blocks of 64 trials. In the first two blocks, the subjects were told to respond to the center instruction by making a saccade to the appropriate target. In the first block, while subjects were learning the instruction, if a saccade was made in the wrong direction, the computer played a distinct tone to indicate an error had been made. After the first block, subjects were not given feedback regarding movement direction; however they made saccades in the wrong direction on only 1.0±0.2% (mean±SEM across subjects) of trials after the first block. Visual observation of the target and the error tone were the main sources of feedback in this task.

Before the start of the 3rd block, the subjects were given new instructions: "For some of the trials, the first instruction may be followed, after a delay, by a

tone [the second instruction]. Occurrence of this tone means that the first instruction has been canceled and replaced. In this case, you should continue fixation." In blocks 3-6, on 25% of the trials after a variable delay period the instruction changed, signaled by a distinct sound. This instruction-change cue was different from the error tone. Success or failure on these trials was determined only by whether subjects responded to the instruction-change cue, and was independent of the saccade direction. Therefore, if the subject followed the first instruction and made a saccade, and the instruction did not change, that trial was a success. If the subject followed the first instruction but the instruction changed, then the trial was a failure, and the error tone, same as that from block 1, was played. If the subject waited, maintaining fixation despite the first instruction, and subsequently the instruction changed, then the trial was a success. The only feedback was the success or failure of the current trial determined only by whether or not the subject made a saccade, indicated by the error tone. Making a saccade in the incorrect direction was not penalized, though this happened rarely. We did not provide scores regarding number of successful trials or any other cumulative feedback. In the final block the instruction did not change, but the subjects were not provided verbal information regarding this fact.

If one were to react only to the first instruction, then one is successful with 75% probability. Waiting for the second instruction improves the probability of

success by 25%. How long would an individual be willing to wait to improve their odds? The variable of interest was the delay period that could be sustained by each individual. The instruction-change delay period started at 200 ms for all subjects. If on the instruction-change trial the subject was successful (i.e., the subject had waited), the instruction-change delay increased by 30 ms, requiring them to wait longer in the future. If on the instruction-change trial the subject failed, the instruction-change delay decreased by 30 ms. Therefore, with this adaptive algorithm we attempted to find how long the subject was willing to wait to acquire the greater odds of success. A formal analysis of this task is provided in Section 5.3.

Each trial was 2.5 seconds in duration. This duration was fixed regardless of events that occurred in that trial. In this way, both the subject that waited a brief period of time for the second instruction, and the subject that waited a long period, experienced the same total experiment time and the same overall rate of movement.

After completion of the task, subjects filled out two questionnaires that are commonly used to measure impulsivity as a psychological profile. These questionnaires were the Barratt Impulsiveness Scale (BIS-11) (Patton et al., 1995), and the I7 Impulsiveness Questionnaire (Eysenck et al., 1985). For the I7 questionnaire, we did not use the components in the empathy category.

Modeling

We considered a model to describe the process of decision making in this task. This model is described in Eq. 5.5 to Eq. 5.9 of the Results section. As a trial began, the model decided at what time it would move using its temporal discount function and expected probability of success, given the expected arrival time of the second instruction. As time progressed (in 1 ms increments), the model truncated its expectation of the probability of the time of the second instruction, updated its desired movement time, and, if the desired movement time was at the current time or sooner, responded to the first instruction, i.e., stopped waiting. Otherwise, the model waited. If the model moved prior to the arrival of the second instruction (a failed trial), the delay of the second instruction was reduced by 30 ms, as in our experiment. Otherwise (a successful trial), the delay was increased by 30 ms. Our model changed its estimate of $\hat{\Delta}$, the expected arrival time of the second instruction, only in trials in which there was a second instruction. Therefore, we simulated our model with 64 trials in which the instruction changed.

Section 5.3: Results

In part 1 of our experiment we asked whether there were consistencies in the saccade velocities of healthy individuals across several amplitudes. Using these velocities, we defined a measure of movement vigor for each subject. In part 2 we asked whether an individual's temporal discounting of reward, as

measured in a decision making task, was a predictor of that individual's movement vigor.

Between-subject differences in movement velocities

Figure 5.2A shows the eye velocity trajectories of two representative subjects during saccades of various amplitudes. Saccade peak velocity and duration increased with amplitude in both subjects, but for any given amplitude subject 4H had peak velocities that were higher than subject 16P. One way to summarize these data is to consider peak velocity as a function of endpoint displacement for each subject. Figure 5.2B provides this data for five representative subjects, measured over 4 days. In this figure, each line represents data from one subject on one day.

We first asked whether there were significant between-subject differences in the saccade peak velocity-amplitude relationship. To determine whether the between-subject differences were statistically robust, we performed a one way repeated measure ANOVA on peak velocity measurements where displacement on each day was the within-subject factor and subject identity was the between-subject factor. We found significant effect of subject identity ($F(4, 15) = 22.5$, $p < 10^{-5}$), and a subject by displacement interaction ($F(60,225) = 45.5$, $p < 10^{-9}$). This indicates that there were highly significant between-subject differences in

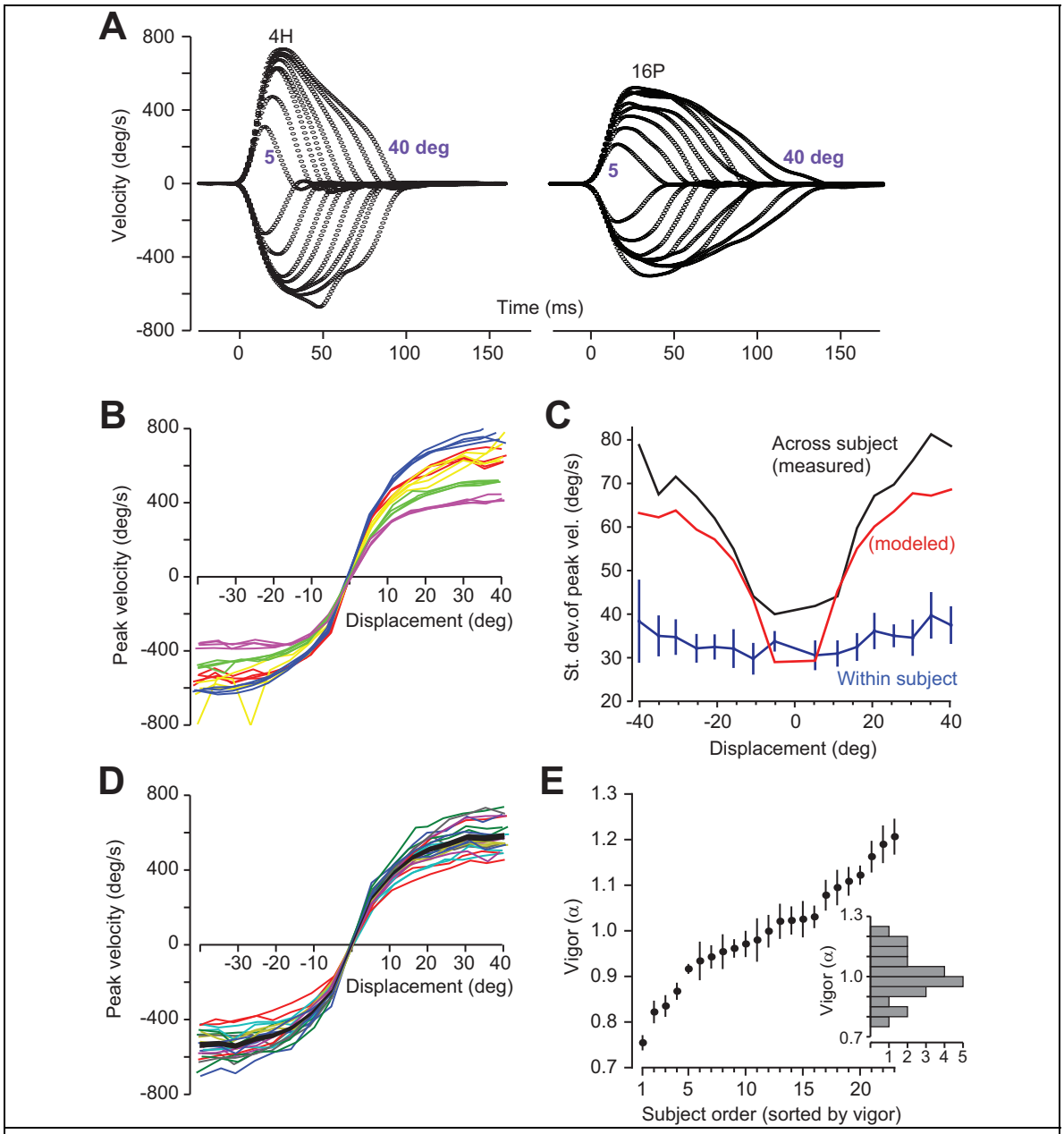

Figure 5.2: Vigor of saccades. A. Average eye velocity traces for horizontal saccades for two representative subjects. Saccades were averaged in 5° increments, centered at 5° to 40°. B. Peak velocity vs. endpoint displacement of the eye during saccades for five subjects on four separate days. Each color is one subject; each line represents a single subject on a single day. Positive represents temporal displacements and velocities. C. Within subject variability of peak velocity, and across subject variability of peak velocity, as a function of displacement. Error bars are SEM. The black line is across-subject measured variability and red line is variability accounted for using the model of Eq. 5.2. Peak velocity was much less variable within a subject than across subjects. D. Peak velocity-displacement relationship for all subjects. Each line represents the data for a single subject. The thick line is the across subject mean±1SEM. E. The distribution of vigor of saccades, across the population of subjects. Vigor of 1 represents the mean of the population. Error bars are 95% confidence intervals in estimating vigor for each subject.

the amplitude-velocity relationship of saccades: day after day, some subjects moved their eyes with reliably higher velocities than others.

To quantify the consistency of velocities within a subject, we asked whether peak velocities are more variable within or across subjects. We measured the standard deviation of peak velocity at a given displacement for each subject. The resulting within-subject distribution is shown in Figure 5.2C (blue line). In comparison, consider the across-subject distribution of peak velocities (Figure 5.2C, black line). The across-subject standard deviation is about twice that of the within subject standard deviation. At all displacements, a t-test produced a significant difference in comparison of the within and between subject measures (all cases $p < 0.0015$). A Bonferroni-Holm correction of the p-values for $m = 16$ family of multiple comparisons demonstrated that the differences remained significant after this correction. Therefore, peak velocity was much less variable within a subject than across subjects. This implies that individuals have a characteristic, trait-like velocity with which they move their eyes.

Movement vigor

The data in Figure 5.2B suggest that the between subject differences in velocity may be summarized by a scaling factor. To see this, consider the velocity-displacement relationship for all subjects, as shown in Figure 5.2D. The

heavy black line is the across-subject mean of the data. Let us label this across-subject mean as $g(x)$, where x is endpoint displacement and $g(x)$ describes the relationship between displacement and average velocity across the population. In Eq. 5.2, we hypothesized that each subject's velocity-displacement relationship is a scaled version of this function, where the scaling factor α_i is our proxy for vigor of saccades for subject i. To test whether Eq. 5.1 is an accurate representation of the data in Figure 5.2D, we found the parameter α_i that in a least-square sense best fitted the amplitude-velocity data for subject i. Next, we used Eq. 5.2 to predict the between-subject standard deviation of saccade velocities as a function of displacement:

$$SD\left[v(x)\right] = SD\left[\alpha\right]\left|g(x)\right| \qquad \text{Eq. 5.3}$$

In Eq. 5.3, $|\ |$ indicates absolute value, and $SD[\ ]$ is the standard deviation operator. We then compared the predicted standard deviation-displacement relationship of Eq. 5.3 (across subject modeled, red line, Figure 5.2C) with the actual relationship (across subject measured, black line, Figure 5.2C). We found that the model and data correlated at r = +0.94, $F(1,14) = 116$, $p < 10^{-8}$. The goodness of fit of our model of vigor for each subject is shown by the confidence intervals in Figure 5.2E, and the resulting distribution of saccade vigor, i.e. parameter α, is shown in the inset of Figure 5.2E. Therefore, it appears that Eq. 5.2 is a reasonable representation of the velocity-displacement data of each

subject and the parameter α can be used as a measure of the vigor of each subjects' eye movements.

We wondered whether differences in vigor are related to differences in endpoint variability. That is, do people who have higher vigor move their eyes with less accuracy? We compared mean standard deviation of saccade endpoints across all target distances with vigor and found that increased vigor did not correspond to more variability ($F(1,21) < 1$, $p = 0.99$). Therefore, accuracy is not a cost that can readily account for between-subject differences in vigor.

If there is a general cost of time for control of movements, then subjects that exhibit a greater vigor (and therefore a greater cost of time) may also exhibit a faster reaction time (RT). We observed a trend in this direction, but the trend was not statistically significant: $r = -0.28$, $F(1,21) = 1.76$, $p = 0.20$. That is, people who had higher vigor did not react faster to a stimulus.

Estimating the temporal discount function

It is possible that between-subject differences in movement vigor are related to between-subject differences in the reward system of the brain (Shadmehr et al., 2010): populations that show increased saccade velocity may also exhibit increased rates of temporal discounting in decision making tasks. For example, rhesus monkeys have saccade velocities that are about twice as fast as humans (Straube and Fuchs, 1997; Chen-Harris et al., 2008). Monkeys have

eye biomechanics that are somewhat different than humans (Fuchs et al., 1988), but once these differences are accounted for, there remains persistent differences in movement vigor (Shadmehr et al., 2010). Intriguingly, monkeys exhibit a greater temporal discount rate: when making a choice between stimuli that promise juice over a range of tens of seconds, thirsty monkeys (Kobayashi and Schultz, 2008; Hwang et al., 2009) exhibit discounts rates that are higher than that of thirsty humans (Jimura et al., 2009).

Let us define temporal discounting as follows:

$$V\left(r,t_0+t\right)=V\left(r,t_0\right)F\left(t\right) \qquad \text{Eq. 5.4}$$

The value of reward at current time t_0, written as $V\left(r,t_0\right)$, is discounted by a function $F\left(t\right)$ to produce value at time t_0+t, with $F\left(0\right)=1$. Suppose subject 1 is given a choice between a small amount of reward now $\left(r,t_0\right)$ and a large amount of reward later $\left(r+R,t_0+t\right)$, and this subject picks the smaller reward. In comparison, subject 2 is given the same choice but picks the larger reward. In this choice, subject 1 is more impulsive, preferring the sooner but smaller reward. Therefore, for subject 1, $V_1\left(r,t_0\right)>V_1\left(r+R,t_0\right)F_1\left(t\right)$, whereas for subject 2, $V_2\left(r,t_0\right)<V_2\left(r+R,t_0\right)F_2\left(t\right)$. If we assume that the two subjects value a given reward equally at the current time, i.e. $V_1\left(r,t_0\right)=V_2\left(r,t_0\right)$, then we infer that the temporal discount function of subject 1 devalues reward more than

subject 2, $F_1(t) < F_2(t)$, which implies that subject 1 is a steeper discounter. According to our hypothesis, subject 1 should generally move with greater velocity than subject 2.

To test this prediction, we designed a task to measure temporal discounting (Figure 5.1B). A critical component of our task was that each choice produced an immediate and real consequence (success or failure), which was experienced before the next choice. As we will see, the consequence of the choice affects the next choice, and this trial-to-trial change in behavior is related to the individual's temporal discount function.

On each trial subjects were given instructions to make a movement. However, on some fraction of trials ε after some time delay Δ there was a second instruction. On trials with only a single instruction, one was successful by following that instruction. On trials with a second instruction, one was successful only after waiting for that instruction. The subjects did not know whether a trial had one or two instructions. Only by waiting the subjects discovered the nature of the trial. The result of each trial, success or failure, reinforced the choice that was made.

The probability of success in a trial increased with waiting. The probability of success, given that the second instruction came at Δ seconds, was described by a logistic function:

$$\Pr(success \mid \Delta) = (1-\varepsilon) + \frac{\varepsilon}{1+\exp\left(-b*\left(t-\tau-\Delta\right)\right)} \qquad \text{Eq. 5.5}$$

In Eq. 5.5, ε is the fraction of trials in which there is a second instruction, τ is the amount of time it takes to respond to the second instruction (i.e., reaction time), t is the time at which the movement takes place, and b reflects the variance in the ability of a subject to reproduce the predicted timing of the second instruction. In this experiment, $\varepsilon = 0.25$. In our simulations $\tau = 0.1$ was used, reflecting the approximate response time (~110 ms, discussed below) to the second instruction. We also used $b = 100\,\text{Hz}$, which corresponds to a standard deviation in the estimate of time of approximately 17 ms. This is similar to estimates of the standard deviation of the ability of subjects to produce a time interval in prior work (17 ms for 400 ms intervals (Ivry and Hazeltine, 1995)). We have plotted Eq. 5.5 via a blue curve in Figure 5.3. The longer one waited before making a movement, the higher the chances of success.

Suppose that from the history of previous trials, the subject estimates the time Δ that the second instruction will come. For example, in Figure 5.3 (top row) the subject expects that the second instruction will come at a time as shown by the red distribution, labeled as $p(\Delta \mid t)$. $\hat{\Delta}$ is the median of this distribution, and this is the best guess, at current time t, regarding when the second instruction will come. If the subject's objective is to maximize probability of

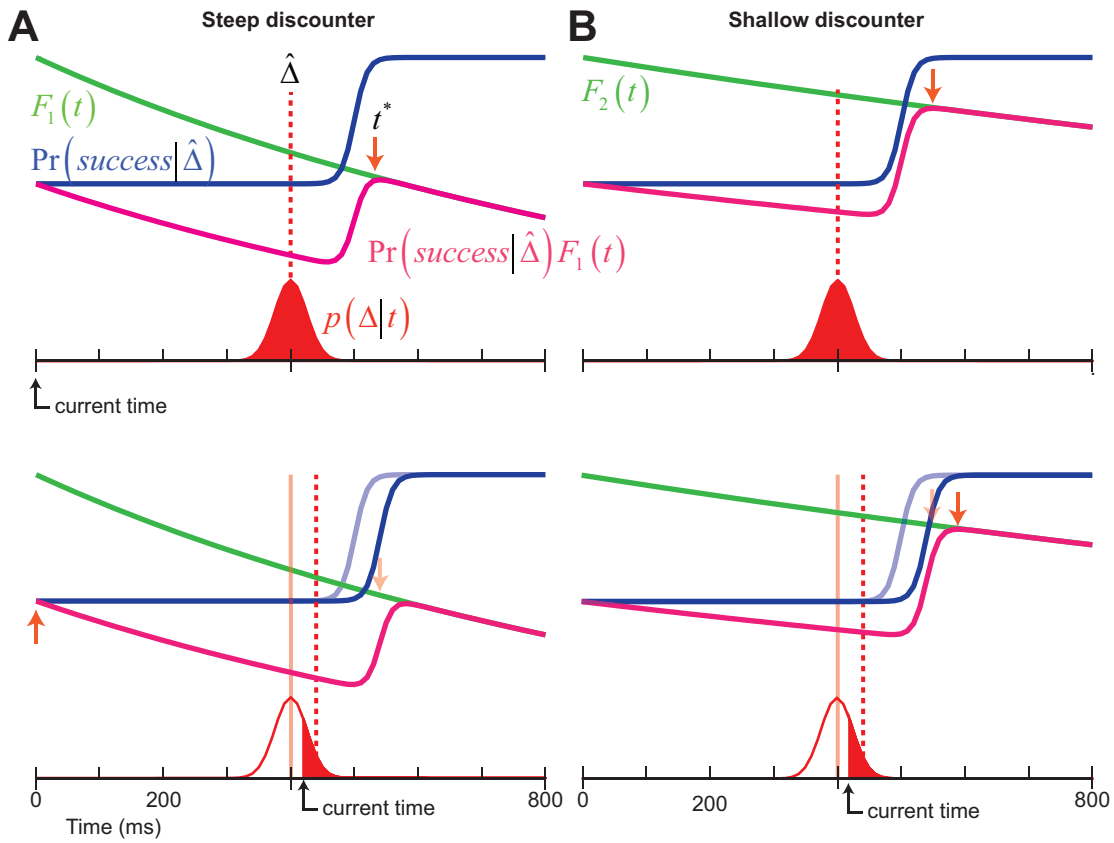

Figure 5.3: Decision making of simulated subjects during each trial. A, A simulated individual with a steep discount rate, and B, a simulated individual with a shallow discount rate. Based on the history of previous trials, and the current time t within the trial, each subject estimates the time $\hat{\Delta}$ the second instruction will come (red, $p(\Delta|t)$). $\hat{\Delta}$ is the median of this distribution (dotted red line). The probability of success (blue) is 0.75, and rises to 1.0 after $\hat{\Delta}$. The hyperbolic temporal discount function (green, $F(t)$) discounts the probability of success to produce the subjective value (pink). For these two subjects, $F_1(t) < F_2(t)$. As the trial starts, $t = 0$ (top row), both subjects estimate the time of the second instruction will be $\hat{\Delta} = 400$ ms. For both subjects it is worthwhile to wait because the peak of the discounted probability $\Pr(success|\hat{\Delta})F(t)$ is in the future. The time of this peak is labeled $t*$. As they wait and time passes, the probability distribution $p(\Delta|t)$ becomes truncated because the time of the second instruction cannot be in the past (second row). As a result, $\hat{\Delta}$ shifts to the right. This change causes a change in the probability of success (blue), as well as the discounted probability of success (red). After 400 ms of waiting, the steep discounter has a discounted success function with a peak that is now in the past. This subject will stop waiting and respond to the first instruction. The shallow discounter, however, still has a discounted probability function with a peak that is in the future. For this subject, the optimum action is still in the future. This subject will continue to wait.

success, then the subject would wait indefinitely on each trial. However, suppose time discounts reward such that:

$$F(t) = \frac{1}{1+\beta t} \qquad \text{Eq. 5.6}$$

Eq. 5.6 is a temporal discount function, representing hyperbolic discounting of reward. This function is shown by the green line in Figure 5.3A. Consider two hypothetical subjects: one who has a steep discount function (large $\beta = 0.58$), as shown in Figure 5.3A, and one who has a shallow discount function, as shown in Figure 5.3B (small $\beta = 0.2$). These two values of β were selected to illustrate the differences in the behavior of subjects with steep and shallow discount functions. Using probability of success and their personal temporal discount function, the two subjects choose the amount of time they are willing to wait so that they maximize the discounted value of success:

$$t^* = \arg\max\left\{\Pr\left(success \mid \hat{\Delta}\right) F(t)\right\} \qquad \text{Eq. 5.7}$$

The discounted value of success is plotted via the pink curve in the top row of Figure 5.3, and the optimum wait time t^* is labeled with an arrow.

Let us illustrate how these two hypothetical subjects would behave on a given trial. As trial *n* starts, i.e., $t = 0$ suppose both subjects estimate that the second instruction will come at $\hat{\Delta}$ = 400 ms, the median of the probability

distribution $p(\Delta \mid t)$. $p(\Delta \mid t)$ was simulated as a Gaussian with mean of 400 ms and standard deviation of 25 ms. This standard deviation is similar to the standard deviation of subjects' perception of 400 ms time intervals in prior work (20.3 ms in Ivry and Hazeltine (1995); ~32 ms in Westheimer (1999)).

For this $\hat{\Delta}$, at $t = 0$ the optimum amount of time to wait is in the future (Eq. 5.7): the discounted value of success $\Pr\left(success \mid \hat{\Delta}\right) F_1(t)$ for both subjects has a maximum that lies in the future. Therefore, both subjects wait. As they wait and time passes, the probability distribution $p(\Delta \mid t)$ becomes truncated because the time of the second instruction cannot be in the past (second row, Figure 5.3). This means that as time passes in the trial and the subject waits, $\hat{\Delta}$ is not constant but becomes larger, reflecting the median of the now truncated $p(\Delta \mid t)$:

$$\hat{\Delta}(t) = Median\left[p(\Delta \mid t) \right] \qquad \text{Eq. 5.8}$$

This change in $\hat{\Delta}$ (dotted red line) causes a change in the probability of success (blue curve), which in turn produces a change in the discounted probability of success (pink curve). The second row of Figure 5.3 shows discounted value of success at t = 410 ms. At this time (i.e., 410 ms into the trial), for the impatient subject (steep discounter) the peak discounted value is no longer in the future, but is in the past (the red arrow is now at t = 0). The impatient subject stops waiting and initiates their movement, responding to the

first instruction. The time of the movement represents the saccade latency of this subject. In contrast, for the person with the shallow discount function (patient discounter), at time t = 410 ms the discounted value of success has a maximum that is still in the future. This person will continue to wait.

In our task, the time of the second instruction, represented by Δ, was adjusted so that it tracked the amount of time each subject was willing to wait. If the second instruction occurred and the subject had waited for it (successful trial), Δ was increased by 30 ms. If the second instruction occurred and the subject had not waited for it (failed trial), Δ was decreased by 30 ms. Using the temporal discount functions shown in Figure 5.3, we simulated behavior of the two hypothetical subjects, as shown in Figure 5.4A and Figure 5.4B. After a trial in which the second instruction occurred, regardless of success or failure the model updated its expectation of the time of this event as follows, where, in our simulations, $\eta = 0.7$:

$$\hat{\Delta}^{(n+1)} = \hat{\Delta}^{(n)} + \eta\left(\Delta^{(n)} - \hat{\Delta}^{(n)}\right) \qquad \text{Eq. 5.9}$$

In the simulated steep discounter, Δ reached a maximum of around 400 ms, whereas in simulated shallow discounter, Δ reached a maximum of around 1400 ms. In Figure 5.4C, we have plotted the asymptotic value of latency for various temporal discount rates. The y-axis of this figure indicates the final saccade latency in the simulated experiment. The model suggests that at the end

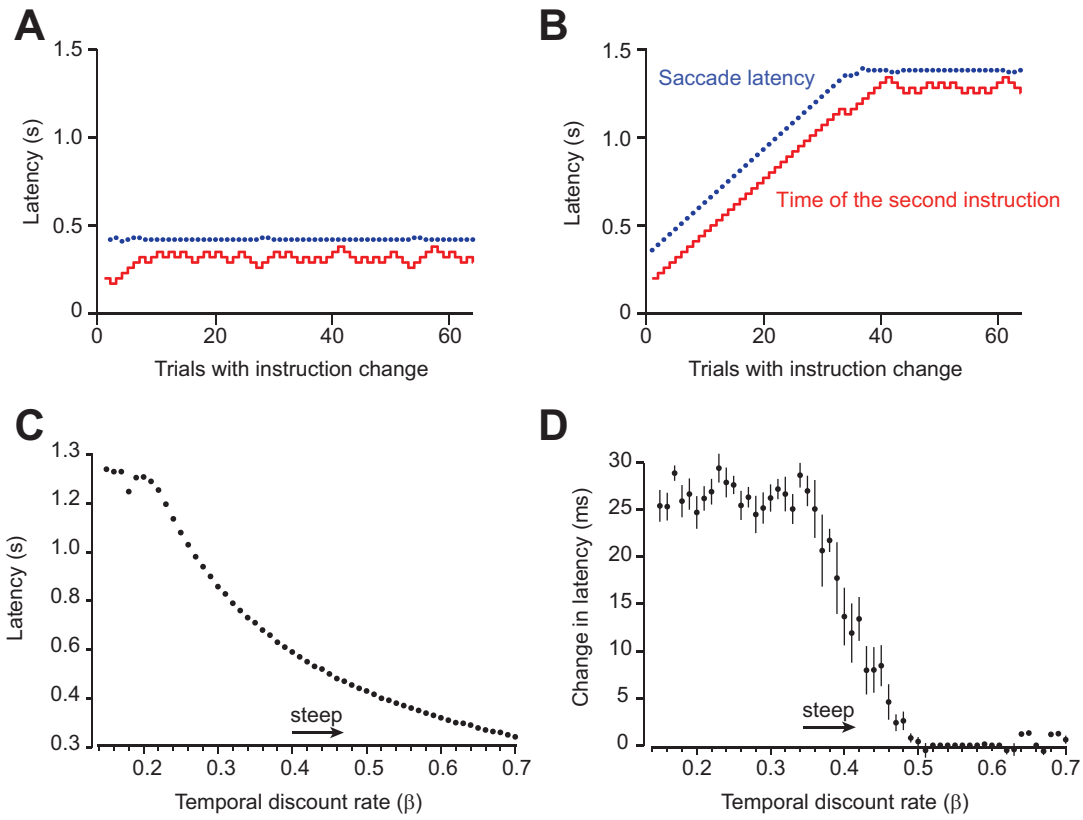

Figure 5.4: Decision making of simulated subjects during the experiment. A, B, Decision making on each trial for simulated subjects (a steep and a shallow discounter) with same discount rates as in Figure 5.3. Saccade latency (blue line) is the time that the simulated subject decided to move. On every trial with a second instruction the delay Δ was adjusted. The steep discounter reached a maximum delay time that was much less than the shallow discounter. **C,** We simulated various discount rates and computed the final latency achieved for each discount function. Steeper discounters are expected to have smaller asymptotic latencies, i.e., wait shorter periods of time. **D,** Average trial-to-trial change in latency following a failed trial (a trial in which the second instruction arrived, but the simulated subject had chosen not to wait) for the first block (16 trials) for 100 simulated subjects. After a failed trial, shallow discounters increase their latency by a larger amount than steep discounters.

of the experiment, a subject that has a shallow temporal discount function will have a longer saccade latency, i.e., will wait longer to respond to the first instruction, than a subject that has a steep temporal discount function.

Whereas the simulations in Figure 5.4C describe the asymptotic behavior of our simulated subjects, the model also makes an interesting prediction with

regard to trial-to-trial change in behavior, in particular near the start of the experiment. Suppose that on trial n, both of our hypothetical subjects predict that the second instruction will come at time $\hat{\Delta}$. Further suppose that in fact, the second instruction comes at a time Δ, later than expected. The two subjects have the same prediction error, and both shift their estimate $\hat{\Delta}$ for the next trial, $n+1$ (Eq. 5.9). Because of the shape of the discount functions, this change in $\hat{\Delta}$ produces a small trial-to-trial change in the latency for the steep discounter, but a larger change in latency for the shallow discounter. To illustrate this idea, we ran our model for various discount functions and focused on the latencies in the first block of trials with a second instruction. We computed how much the simulated subjects changed their latency in response to a trial in which the second instruction occurred but they did not wait for it. That is, we computed the change in behavior in response to a failed trial. The results are shown in Figure 5.4D. The model predicted that subjects with shallow temporal discount functions should respond to a failed trial with relatively large change in latency, whereas steep discounters should show a small change in latency.

Our simulations also illustrate that the change in latency in response to a failed trial is a steeper function of temporal discounting than asymptotic latency. For example, the ability of the model to distinguish a 0.35 discounter from a 0.45 discounter is about 4 times better with the change in latency measure as compared to the asymptotic latency measure. This implies that for two people

who are near the mean of the population, small differences in temporal discount rates will be more easily observed in terms of change in latency as compared to asymptotic latency.

In summary, the results of the decision making task provide two proxies for the rate of temporal discounting: trial-to-trial change in latency following a failed trial as expressed early in the experiment, and asymptotic latency as expressed late in the experiment.

Relationship between vigor and willingness to wait

To verify the validity of our vigor model, we first asked to what extent the vigor estimate for a subject in Exp. 1 was a predictor of their saccade peak velocities in Exp. 2 (the experiments were conducted on separate days). We computed the mean saccade velocities in the first two blocks of Exp. 2 (i.e. baseline blocks) and found that vigor in Exp. 1 was strongly correlated with velocities recorded in Exp. 2 ($r = +0.89$, $F(1,21) = 81.9$, $p < 10^{-7}$).

Saccade latencies of two subjects in Exp. 2 are shown in Figure 5.5A. These subjects are the same ones for which we displayed saccade velocities in Figure 5.2A. In the first two blocks, the probability of a second instruction was zero. In the subsequent four blocks this probability increased to 0.25. At the start of the third block, Δ, representing the delay to the second instruction, was

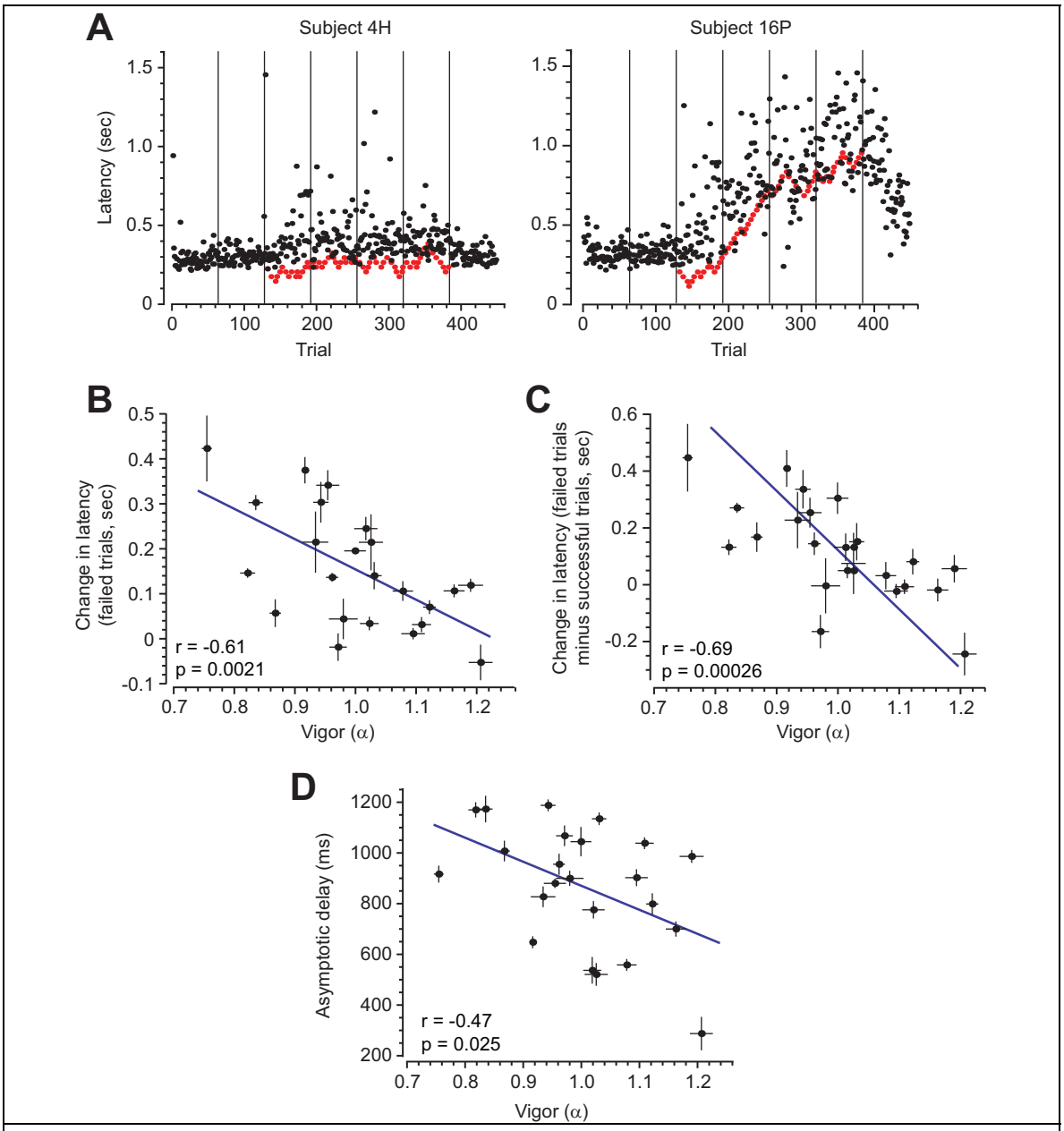

Figure 5.5: Willingness to wait and movement vigor. A. Latency of saccades (black) and delay of the second instruction (red) for two subjects (saccade velocities of these subjects, as measured in part 1 of the experiment, are shown in Figure 5.2A). Vertical lines denote breaks between blocks. B. Trial-to-trial change in saccade latency in response to an error-trial, i.e., a trial in which a second instruction occurred but the subject chose not to wait. The latencies were measured in the third block of the experiment (first block in which the second instruction occurred). C. Trial-to-trial change in saccade latency in response to an error-trial, normalized to the change when subjects successfully waited for the second instruction. Latencies were measured in the third block of the experiment. D. Relationship between asymptotic delay of the second instruction and saccade vigor. Delays were measured during the last half of the final block of trials in which there was a second instruction (block 6). Data are mean±SEM for latency and delay and mean±SD for vigor. Blue lines are results of linear regression.

200 ms. By the end of the 6th block, Δ had increased to around 250 ms for subject 4H, whereas it had increased to around 900 ms for subject 16P.

Our model suggested two proxies for the rate of temporal discount function: change in latency early in the experiment following a failed trial, and asymptotic latency late in the experiment. For change in latency, we focused on the first instruction-change block (third block overall) because latency changes became smaller as the experiment proceeded (paired t-test, latency change following a failed trial, first block with a second instruction vs. last block with a second instruction, $t = 2.62$, $p = 0.016$) and because at this point in the experiment, subjects have similar instruction-change delays (vigor vs. instruction-change cue delay, $r = -0.31$, $p = 0.15$). For every failed trial (in the first block in which the second instruction occurred), we computed the change in latency from the single-instruction trial before to the single-instruction trial after. We found that subjects who displayed more vigor in their saccades tended to have a small change in latency, as shown in Figure 5.5B ($r = -0.61$, $F(1,21) = 12.3$, $p = 0.002$). This relationship was maintained when we normalized the change in latency with respect to successful trials (trials in which the second instruction occurred and the subject waited). We computed the change in latency following a failed trial and subtracted from it the change in latency following a successful trial. The subjects who displayed more vigor in their saccades had a strong tendency to exhibit a small change in this normalized measure of latency, as

shown in Figure 5.5C ($r = -0.69$, $f(1,21) = 19.2$, $p = 0.0003$). That is, people with high vigor were less willing to increase their latency following a failed trial, less willing to wait longer to improve their odds of success.

A second variable of interest was the asymptotic value of the instruction delay. We quantified this by taking the mean delay during the last half of the final block in which the second instruction occurred. Across our sample of volunteers, we found a negative correlation between the instruction delay and movement vigor: people who had higher saccade vigor tended to achieve a short instruction delay (Figure 5.5D, $r = -0.47$, $F(1,21) = 5.8$, $p = 0.025$). A similar result was found when we compared vigor with the maximum instruction delay achieved by each subject ($r = -0.49$, $F(1,21) = 6.74$, $p = 0.017$).

Individual differences in valuations of immediate reward

In our model of temporal discounting, we assumed that the value of a delayed reward depended on its immediate value, multiplied by a function that discounted this value as a function of time (Eq. 5.4). We made the assumption that two subjects differed in the rate of temporal discounting, but not the immediate value. In other words, we assumed that subjects did not differ in how they valued success in a given trial in which they did not have to wait (V in Eq. 5.4). Is there a way to verify this assumption?

Generally, if one action is valued more than another, people (Milstein and Dorris, 2007) and animals (Tachibana and Hikosaka, 2012; Kim and Hikosaka, 2013) will react with a shorter latency or arrive at the target earlier in the more valuable scenario. Therefore, differences in the value of success may produce between subject differences in reaction time (RT) or target acquisition time (RT + movement duration) in the baseline period, i.e., in the block in which there was no instruction-change. While there were wide differences between people in the baseline blocks, such differences did not correlate with vigor (RT vs. vigor: $r = -0.13$, $F(1,21) = 0.36$, $p = 0.55$; target acquisition time vs. vigor: $r = -0.23$, $p = 0.27$). That is, people with greater vigor were not faster in reacting to the first instruction.

There was no relationship between asymptotic delay and baseline target acquisition time ($r = 0.25$, $p = 0.25$) nor between the change in latency and baseline acquisition time (failed trials: $r = 0.06$, $p = 0.78$; failed minus successful trials: $r = 0.07$, $p = 0.75$). There was also no correlation between baseline reaction time and our measures of temporal discounting (asymptotic delay vs. RT, $r = 0.23$, $p = 0.29$; change in latency, failed trials vs. RT: $r = 0.08$, $p = 0.83$; failed minus successful trials vs. RT: $r = 0.05$, $p = 0.83$). People who had larger asymptotic delays or larger changes in trial to trial latency were not faster in reacting to the first instruction.

Prior work has recently demonstrated that differences in the implicit value of a stimulus may be reflected in the peak velocity of saccades to that stimulus (Xu-Wilson et al., 2009). However, in that work the change in the peak velocity for stimuli of differing value was ~5 deg/sec for a 15° saccade, more than an order of magnitude smaller than the differences in peak velocity between subjects in the work in this chapter. Accordingly, the differences in vigor between subjects are unlikely to be driven by differences in stimulus value.

A difference in the value of the stimuli could also be reflected in the rate in which subjects followed the directional cue given by the first instruction. After the first block, subjects made saccades in the wrong direction only 1.0±0.2% (mean±SEM across subjects) of the time. The accuracy of subjects was not correlated with the vigor of their movements ($r = -0.01$, $p = 0.98$). Overall, these analyses suggest that there was no systematic difference in the way subjects valued the stimuli.

Reaction to the instruction-change cue

Our task for measuring temporal discounting was similar to the Stop Signal Reaction Time (SSRT) task, a task in which subjects are provided with a 'go' cue, which is occasionally followed with a 'stop' cue. The objective of the SSRT task is to measure how long it takes after the occurrence of the stop signal for the subject to abort their planned movement (this latency is called SSRT). An

important difference in SSRT versus our task is that in SSRT, subjects are instructed to respond to the go-cue as quickly as possible and not delay their response in order to await the stop cue. In our task, the subjects were told that occasionally there will be a second instruction. They were allowed to wait as long as they wished to respond to the first instruction. Despite this difference, we thought it worthwhile to analyze our data to quantify how behavior was affected in trials in which instruction-change occurred.

We began by asking whether saccade kinematics were different in the instruction-change trials (i.e., the failed trials). Our thought was that (partial) inhibition of a planned movement may produce a reduction in its amplitude. Indeed, subjects made significantly smaller saccades in instruction-change trials (19.8±0.3°, mean±SEM) as compared to no-change trials (20.6±0.1°; paired t-test, $p < 0.003$). We considered the amplitude of saccades made in those trials as a function of when the saccade was made relative to the second cue. We found that if a saccade occurred 40 ms or later after the instruction-change cue, its amplitude was reduced as compared to no-change trials (18.9±0.5°, mean±SEM across subjects; paired t-test, $p < 0.0005$). However, saccades made prior to 40 ms after the instruction-change cue showed no amplitude differences (20.6±0.2°, paired t-test $p = 0.62$). The saccades were on average 67±1.3 ms in duration. Therefore, it took a minimum of ~110 ms after the instruction-change cue for the

brain to alter the ongoing motor commands. This value provides an objective estimate of the lower bound on the SSRT in our task.

To estimate SSRT for each subject, we used the approach suggested by Eagle et al. (2008) for experiments in which the timing of the second instruction is adjusted via an adaptive 'staircase' procedure: we subtracted the median of the latency for the second instruction from the median of the reaction time in the trials without a second instruction. We found that on average, SSRT was 120.3±11.3 ms (population mean±SEM), which agrees well with our independent measure using saccade kinematics (lower bound of 110 ms). A within subject comparison of SSRT and vigor did not result in a significant correlation ($r = +0.23$, $F(1,21) = 1.18$, $p = 0.29$). People who require a long time to inhibit a planned action (manifested in long SSRT) are thought to be more impulsive. Therefore, the positive value of the correlation, though not significant, is in line with our general framework. Our task, however, was not designed to measure SSRT.

Psychological profile

A commonly used method to assess decision making characteristics of individuals is via questionnaires that measure impulsiveness. These questionnaires estimate personality traits by determining the response to queries such as "do you often buy things on impulse", "do you mostly speak before

thinking things out" etc. Our subjects filled out two commonly used questionnaires, termed BIS and I7. Higher scores in these questionnaires suggest a psychological profile for impulsiveness. In our subjects, the score in one questionnaire was strongly corrected with the score in the other ($r = +0.65$, $F(1,21) = 15.5$, $p < 0.001$). However, impulsivity as measured by these questionnaires was never a good predictor of movement vigor (I7 impulsivity sub-score vs. vigor, $r = +0.20$, $p = 0.35$; BIS vs. vigor, $r = +0.17$, $p = 0.45$), nor of the asymptotic instruction delay (I7 impulsivity sub-score vs. delay, $r = -0.10$, $p = 0.66$; BIS vs. delay, $r = -0.01$, $p = 0.95$). The positive correlation values for vigor, and the negative correlation values for decay, indicate that in general people who score slightly more impulsive on the questionnaires tended to have higher vigor and slightly shorter delays, though this tendency was not significant.

Section 5.4: Discussion

We found consistent differences among healthy people in the speed with which they moved their eyes during a saccade. We quantified this via a measure of vigor that summarized the relationship between saccade amplitude and peak velocity. Vigor differed by as much as 50% between subjects, but was highly consistent within a subject. Why do some people make movement quickly, when others make slower movements? We hypothesized that differences in vigor may

be partly due to differences in how the brain discounts reward as a function of time.

To measure temporal discounting, we considered a task in which subjects received instructions to perform an action, but improved their odds of success if they waited for a second instruction. We found that people with high vigor were less willing to increase their latency following a trial in which they failed, suggesting a higher temporal discount rate. This measure of temporal discounting in the decision making task accounted for 48% of the between subject variance in vigor. Movement speed is, in part, related to the willingness of subjects to wait to acquire higher reward – the rate at which they discount reward in time.

To what extent differences in vigor can be explained with differences in biomechanics? In a recent study, the eyes, orbit, and extraocular muscles of healthy volunteers were imaged using MRI (Peng et al., 2012). That study concluded that the measured parameters (including muscle volume and cross-sectional area) could not account for the between-subject differences in saccade velocity. Biomechanics of the eyes are critical in describing consequences of motor commands, affecting potential costs of movements in terms of effort and variability. In our population, endpoint variability of saccades was unrelated to vigor, i.e., people with high vigor were not more variable, as might be expected if

all subjects exhibited the same signal-dependent noise function. We explore the variability in the signal dependent noise of subjects in Chapter 6.

To measure temporal discounting, we designed a task in which choices had consequences (success or failure) that acted as operant reinforcers before the next choice was made. Indeed, we relied on the fact that the reinforcer caused a change in behavior from one trial to the next, and the magnitude of this change was a signature of the temporal discount function. In contrast, most experiments that measure temporal discounting in humans rely on non-operant reinforcers in which people make choices between dollar amounts, but the consequences are either hypothetical, or are realized only after the experiment is over (because the delays are in days or weeks). While both types of experiments produce measures of temporal discounting, they produce inconsistent results in the same person (Hyten et al., 1994), and produce discount rates that differ by many orders of magnitude (Navarick, 2004). The operant approach is the principal method of measuring discounting in non-human primates, which guided our design here.

Our task is similar to the Stop Signal Reaction Time (SSRT) task. In SSRT, subjects are provided with a stimulus that instructs a movement, but are told to not delay their response to this instruction. In SSRT, the objective is to measure how quickly subjects can stop their planned movement in the case that a 'stop' instruction appears. In our task, the subjects were allowed to wait as long as they wished. When the instruction changed, we estimated 110 ms as the lower

bound for the time it took the brain to process the new instruction and alter the saccade. People who need a longer time to inhibit their movements exhibit impulsivity (Verbruggen and Logan, 2008), and in our sample such people tended to exhibit greater vigor, though the effect was not statistically significant.

The key variables in our task were the change in latency after an unsuccessful trial, and the asymptotic latency, both of which we found to negatively correlate with vigor. A limitation of our model, however, is that the change in latency that it predicted for a given discount function was a scaled version of the actually observed values. It is unclear to us where this limitation arises from. It may indicate an asymmetry in valuation of success vs. failure.

The mathematical framework of optimal control predicts a link between vigor and temporal discounting by suggesting that before a movement can be generated, there needs to be an evaluation of the reward that is expected at the end of the movement, discounted by the time it takes to complete that movement (Shadmehr et al., 2010). Of course, it is possible that there are two separate temporal discounting systems for control of movements and decision making, as the two have vastly different timescales. However, if the basis of both forms of temporal discounting is to maximize discounted rate of reward (Haith et al., 2012), then from a theoretical perspective there is justification for the idea that there is a single temporal discounting system that affects control of movements as well as decision making. From an evolutionary perspective, control of

movements may have required temporal discounting, which in turn was generalized to control of decisions.

Reaction time and vigor

It is possible that during the RT period, the brain is solving the problem of 'what is the best action that I can perform?', while during the movement, the brain is solving the problem of 'how do I perform this action?' Indeed, when there are two possible actions, during the RT period there is competition between the two actions: the brain accumulates evidence for each action, and the action that reaches a threshold first is selected (Gold and Shadlen, 2002). A person that has a high cost of time should, in principle, have a lower threshold, selecting actions earlier and with less evidence. If there are differences in cost of time between people, and if these costs generalize between action selection and action execution, then there should be a negative correlation between vigor and RT.

Impulsivity and vigor

As people wait for an expected reward, activity (as measured by fMRI) in the ventral striatum and ventromedial PFC rises, and this rise has a steeper slope for people who have a steeper temporal discount function (Jimura et al., 2013). Impulsivity is a psychological trait that is often measured via questionnaires. Impulsive people show diminished midbrain D2/D3 auto-receptor availability, which results in increased dopamine release in the striatum (Buckholtz et al.,

2010). In our sample of subjects, there was a positive but not significant correlation between survey-based measures of impulsivity and movement vigor. We suspect that the reason for this is that impulsivity is a complex trait that involves interactions between the basal ganglia and the frontal lobe. For example, in humans the temporally discounted value of reward is correlated with activation in the medial prefrontal cortex (Jimura et al., 2013), in addition to the dorsal and ventral striatum (Kable and Glimcher, 2007; Pine et al., 2009). The cost of time as reflected in saccade vigor may be due to the control that the basal ganglia imposes on the superior colliculus, which in turn is affected by dopamine, whereas the cost of time as reflected in decision making is a more complex process that involves interactions between the basal ganglia and the cerebral cortex.

Chapter 6: Optimal control of saccades in ataxia-telangiectasia[4]

Section 6.1: Motivation

In clinical medicine, when patients produce behaviors or experience sensations that differ from those of a healthy person, we often label the differences as symptoms, or signs, of a disease. For example, the heart normally beats in a steady rhythm, but when the heart rate is abnormally high (tachycardia) we can label the deviation as problematic. Deviations from the norm are not necessarily impairments, however, as they might alternatively reflect a compensatory response to another primary deficit. For example, tachycardia at rest may be the consequence of an arrhythmia to be treated with rate-slowing beta blockers, but may instead be a desirable compensation to maintain perfusion when there is blood loss. A symptom can thus be a primary disorder or a compensatory response to another underlying deficit. Correctly distinguishing between these possibilities is important as symptomatic treatment may be helpful or harmful depending on the underlying cause of the symptom.

Whereas symptoms of neurological disorders are often easy to observe, in some cases it is difficult to distinguish those features that reflect the primary

[4] The work in this chapter is in submission: Vaswani PA, Crawford TO, Wright JM, Lederman HM, Shadmehr R. Dissociating motor symptoms as primary deficits vs. compensatory adaptation: Ataxia-telangiectasia

deficit from those that are compensatory responses. Here, we use a mathematical formulation of the problem of motor control to distinguish the nature of one behavioral symptom – deficit or helpful compensation – of a neurological disease in which that symptom was previously of unknown provenance.

We studied the saccadic eye movements of people with ataxia-telangiectasia (A-T), a rare autosomal recessive disorder in which affected individuals exhibit neuro-degeneration, ocular and cutaneous telangiectasia, and immunodeficiency (Boder and Sedgwick, 1958; Perlman et al., 2003, 2012). Individuals with A-T display progressive ataxia but have intact sensation and normal intelligence (Perlman et al., 2012). Among their neurologic symptoms are impairments in eye movement control – including saccadic pursuit, reduced vestibulo-ocular reflex gain, impaired optokinetic reflex, difficulty with gaze holding, and hypometric saccades – though they have normal visual acuity (Lewis et al., 1999; Farr et al., 2002). Are all of these symptoms primary deficits, or are some (if any) compensatory adaptations to another, more fundamental, impairment?

To approach this problem, we measured saccadic eye movements of people with A-T and noted hypometria and increased variability. Previous work had also observed the symptom of saccadic hypometria but conjectured that it was a manifestation of a deficit in the brainstem oculomotor circuitry (Lewis et

al., 1999). Here, we considered the possibility that this symptom was not a primary deficit, but was instead a compensatory adaptation. We began with a mathematical model of motor control that accounts for saccadic eye movements of healthy people (Shadmehr et al., 2010; Haith et al., 2012; Choi et al., 2014). In this framework, there is a cost that defines the goodness of the movement, providing a metric that one can use to compare various ways to move the eyes toward a target. We extended this mathematical framework from generating a single movement to generating a series of movements and found that, in the presence of increased endpoint variability in saccade execution, the best response to capture a visual stimulus was not a single saccade, but a series of saccades. Thus, the theory as applied to A-T suggests that the symptom of hypometria is a compensatory adaptive response. Hypometric saccades are intentionally selected to compensate for the primary deficit of increased variability, in order to move the eyes with minimal endpoint error in a reasonable time.

Section 6.2: Methods

Paradigm

We measured the saccadic eye movements of individuals with A-T and control children and adults without neurologic problems. Ten subjects with A-T (4 to 30 (19.5±9.1, SD) years old, 4 females) and 45 control subjects without known neurologic disease (4 to 33 (16.3±7.2, SD) years old, 20 females)

participated in this study. All subjects provided informed consent and all experiments were approved by the Johns Hopkins Institutional Review Board.

Subjects sat comfortably in front of a computer screen with their chin and forehead supported by the apparatus. A parent or experimenter also gently held the head of some A-T patients. The left eye was covered with an eye patch, and the gaze of the right eye was recorded at 500 or 1000 Hz using an Eyelink 1000 (SR Research) infrared recording system.

Subjects were asked to make gaze shifts of 6 amplitudes (10°, 15°, 20°, 25°, 30°, and 40°) to visual targets (1.5° diameter) presented on the screen. In each block, targets were sequentially presented to the left and right of center, so that subjects shifted their gaze leftwards and rightwards 10 times (Figure 6.1A). Some subjects also repeated the experiment on the same or next day; data from separate sessions was pooled for each subject.

Saccades were detected when velocity was 20 deg/sec for more than 4 ms and acceleration was greater than 8000 deg/sec^2. Saccades with blinks, duration less than 10 ms or more than 200 ms, or peak velocity greater than 1500 deg/sec were excluded. We selected trials where the gaze was within 4° of the previous target at the time of target relocation to eliminate anticipatory saccades. Leftward and rightward trials of the same amplitude were grouped together. To count the number of saccades made when shifting gaze, we counted saccades that moved the eye at least 1° towards the target, and ended at least 1° closer to

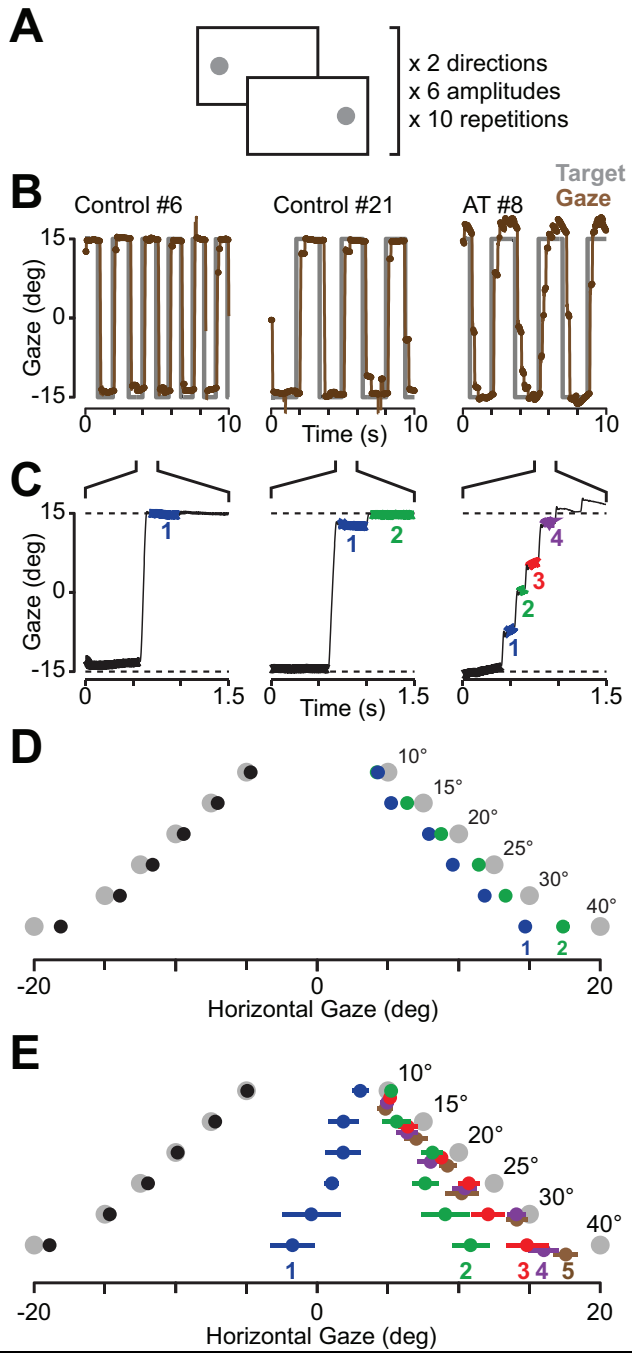

Figure 6.1: Saccades of A-T and control subjects**. A,** Paradigm. Targets were placed between 10° and 40° apart on the horizontal axis. **B**, Representative control (left, middle) and A-T subjects (right). Participants shifted their gaze (brown) to the targets (gray). Thick lines are the periods when the eye was relatively stationary. **C,** Example gaze shifts from these three subjects, expanded from B. The starting eye position (black) and endpoints of each saccade (colors) are shown. **D,** Saccade endpoints of the group of Control and **E**, A-T subjects. The starting eye position and endpoints of each saccade are colored as in C. Targets (gray) are to scale. Data are mean±SEM across subjects. SEM bars for the control group are often too small to be observed and are contained within the points indicating the mean.

the target than the endpoint of the prior saccade. These criteria removed from our count the saccades during fixation of the target that constitute the fast phase of nystagmus.

Optimal Feedback Controller

We used an optimal feedback control model to predict the number and gain of saccades. Subjects can make a saccade with a desired amplitude $u^{(k)}$, which moves the eye from its prior position $x^{(k)}$ to its next position $x^{(k+1)}$, where x represents position with respect to target (Eq. 6.1). Movements are, however, corrupted by noise. In our model, the variance of the consequence of the saccade increased with the amplitude of the desired saccade. That is, noise was signal dependent, where *c* reflects the rate of increase of the variance:

$$\begin{aligned} x^{(k+1)} &= x^{(k)} + u^{(k)}\left(1 + c\phi^{(k)}\right) \\ \phi^{(k)} &\sim N(0,1) \end{aligned} \qquad \text{Eq. 6.1}$$

The controller finds the optimal series of saccades by minimizing the total cost J of producing those saccades. We imposed a quadratic accuracy cost J_X (Harris and Wolpert, 1998), a hyperbolic time cost J_T (Shadmehr et al., 2010; Shadmehr and Mussa-Ivaldi, 2012), and a quadratic effort cost J_U (Todorov and Jordan, 2002).

$$
\begin{aligned}
J &= J_X + J_T + J_U \\
J_X &= \left(x^{(n)}\right)^2 \\
J_T &= \alpha\left(1-\left(1+\beta T\right)^{-1}\right) \\
J_U &= \rho\sum_{k=0}^{n-1}\left(u^{(k)}\right)^2
\end{aligned}
\qquad \text{Eq. 6.2}
$$

Where $x^{(n)}$ is the position after the last (n-th) saccade relative to the target position, T is the total duration of the series of saccades (Eq. 6.3), and α, β, and ρ are constants. In this framework, missing the target, taking a longer time to achieve the target, and making larger saccades are each penalized. The total duration of the series of saccades T was calculated using the approximately linear relationship between saccade amplitude and saccade duration (Collewijn et al., 1988), and a fixed inter-saccadic interval, denoted by variable s:

$$
T = \left(n-1\right)s + \sum_{k=0}^{n-1}\left(a + bu^{(k)}\right) \qquad \text{Eq. 6.3}
$$

We used $\alpha = 15$, $\beta = 0.8$ s-1, $\rho = 0.002$ for all subjects. However the qualitative predictions of the model were not sensitive to modest changes in parameter values. The saccade amplitude-duration relationship ($a = 24$ ms, $b = 2.5$ ms/°), intersaccadic interval ($s = 400$ ms), and signal dependent noise constant for control ($c = 0.13$) and A-T subjects ($c = 0.51$) were estimated from the data (see Section 6.3).

We computed the optimal policy using the Bellman equation (Todorov, 2005; Shadmehr and Mussa-Ivaldi, 2012), approximating the hyperbolic time cost

using a second order Taylor series approximation (for derivation, see Appendix A). The optimal policy given the cost function is a gain and an offset on the remaining distance to the target.

$$u^{(k)*} = -G^{(k)}x^{(k)} - H^{(k)}$$

To find the optimal number of saccades for each target distance, we computed the cost of making n = 1 to 10 saccades under the optimal policy, and identified which series of saccades resulted in the minimum cost.

Simulations

We used the optimal feedback controller described above to find the predicted number, gain, and endpoint of saccades made to each target. Because gaze was not necessarily precisely at the center of the previous target position when the target jumped, we used the actual distribution of target distances in our simulations.

We first computed the distribution, in 1° bins, of the initial target distance for each subject and each target. Then, we found the mean distribution of target distances across all subjects to each target. Using this distribution, we computed the number of saccades, and mean gain and mean endpoint of each saccade to each target (Figure 6.6C and Figure 6.6D) or pooled across targets (Figure 6.6E).

Statistics

All statistical analyses were conducted using Matlab (R2012a, The Mathworks) or SPSS (V21, IBM). All data are mean±SEM unless noted otherwise. Student's t-test was used to compare groups. When comparing saccade dynamics, because group sizes were unequal, we used a general linear model (SPSS) with repeated measure of amplitude, across subject factor of group, and an autoregressive heterogeneous covariance matrix to test for effects of group and group x amplitude interaction on each parameter of saccade dynamics.

Section 6.3: Results

We asked control subjects and subjects with A-T to make saccadic eye movements to small targets (1.5° diameter) presented on a screen, and recorded the gaze of their right eye (Figure 6.1A). Data from two representative control (left, middle) subjects and an A-T subject (right) presented with targets at 30° are shown in Figures 1B and 1C. In response to the 30° target, the first of these control subjects often made a single saccade, resulting in a horizontal error of -1.6±0.4° after his first eye movement (mean±SEM, negative error indicates hypometria). In 50% of trials, after a brief delay, he made a second, corrective, saccade bringing the eye to -0.7±0.3° from the target. The second control subject did not make saccades directly to the target. Instead, his first saccade was consistently shorter, with a horizontal error of -5.7±2.5°. In 53% of trials, he

made a second saccade, bringing him to -1.3±1.2° of the target. The A-T subject depicted also did not make saccades directly to the target, but instead made a series of saccades, typically converging on the target after 2-6 movements. This A-T subject had a horizontal error of -17.0±1.2° after his first saccade to this target. He made 3.3 movements on average; only 3% of trials consisted of a single saccade.

To characterize this behavior in the two groups, we considered the endpoint of each saccade during the trial. With their first saccade, control subjects moved on average 88.6±0.9% of the distance to the target (Figure 6.1D, data pooled across all targets, mean±SEM across the group). With their second saccade, they moved 94.2±0.6% of the total distance to the target, and were -1.3±0.1° from its center. A-T subjects, in contrast, moved only 60.7±4.4% of the distance to the target with their first saccade (Figure 6.1E, t-test, A-T vs. control, $t(53) = 9.89$, $p = 1E\text{-}13$). This was not due to an inability to achieve the target eventually: after 4 saccades, A-T subjects were -1.6±0.4° from the center of the target (t-test, vs. control 2nd saccade, $t(53) = 0.83$, $p = 0.4$), and had moved 94.0±2.1% of the target distance (t-test, vs. control 2nd saccade, $t(53) = 0.15$, $p = 0.9$). Therefore, A-T subjects were capable of capturing the target with normal accuracy, but preferred to produce a series of saccades rather than a single saccade.

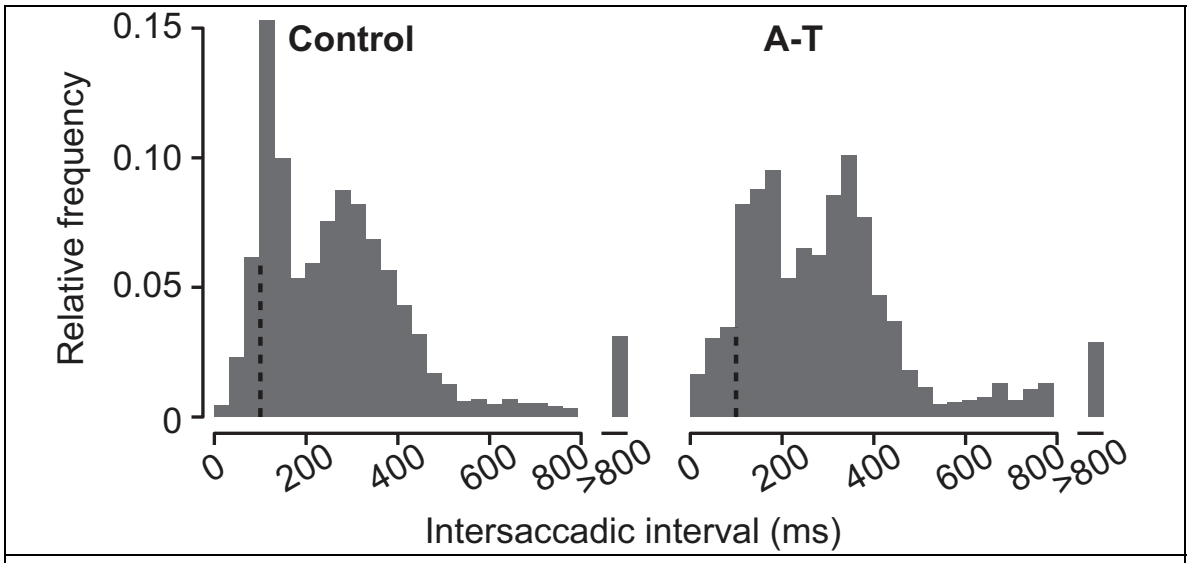

Figure 6.2: Intersaccadic interval (ISI) between the end of the first saccade and the beginning of the second saccade made in a trial, pooled across subjects. The ISI is rarely less than 100 ms (dashed line), suggesting visual feedback was available when making saccades.

We next considered the inter-saccade interval (ISI), the duration of time between the completion of the first saccade and the start of the second saccade in trials in which there were 2 or more saccades. In control subjects, this ISI was 287±8 ms (mean±SEM), with a distribution shown in Figure 6.2. Similarly, the ISI for the A-T subjects was 318±19 ms. We found no statistically significant difference in the ISI of the two groups ($t(53) = 1.7$, $p = 0.1$). Fischer and Ramsperger (1984) have demonstrated that healthy subjects can observe a target and initiate a saccade in approximately 100 ms. Both control (9.0±1.2% of trials) and A-T subjects (8.5±2.5% of trials) rarely made subsequent saccades with an ISI less than this interval (control vs. A-T, $t(53) = -0.2$, $p = 0.9$). This would suggest that the sequences of saccades made by the control and A-T subjects were benefiting from visual feedback regarding target position.

If A-T subjects can eventually capture the target with accuracy comparable to control subjects, why do they make a series of saccades, instead of moving directly to (or near to) the target? Lewis et al. (1999) proposed one hypothesis. They observed that the peak velocities of the saccades of some A-T subjects were significantly higher than those of a control population, and suggested that saccades were programmed to move the eyes to the target, but were prematurely terminated. That is, they hypothesized that the movement was planned to the target, but was quenched mid-flight. They speculated that erroneous activity of fixation neurons in the superior colliculus was truncating saccades. However, our recent work has uncovered substantial between-subject differences in saccade peak velocities (as much as 50%) within a control population (Choi et al., 2014). This raises the possibility that the between-subject differences in peak velocity observed by Lewis et al. (1999) may have been due to natural inter-individual differences. The alternate hypothesis is that the brain of individuals with A-T purposefully programmed a grossly hypometric first saccade, with the aim of arriving at the target after a series of saccades. To test the relative merits of these competing hypotheses, we examined the saccade dynamics of A-T and control subjects.

Saccade dynamics

The hypothesis that saccades are prematurely truncated would predict that the velocity profiles should appear abnormal. The velocity profiles of

saccade of two example subjects, one control and one A-T, are shown in Figure 6.3A. These subjects were selected because they had similar peak velocities for saccades of comparable amplitudes. Notably, the velocity profiles in both subjects were smooth, appeared bell-shaped with a slight skew for large saccades, and displayed no evidence of premature termination.

To further characterize the saccade dynamics, we considered the relationship between the amplitude, duration, and peak velocity of saccades. There were no differences in the amplitude-duration relationship between the two groups (Figure 6.3B, main effect of group, $F(1) = 0.2$, $p = 0.7$; group x amplitude interaction, $F(5) = 2.1$, $p = 0.07$). While A-T subjects did, on average, have somewhat greater peak velocity than the control group (Figure 6.3C, main effect of group, $F(1) = 7.2$, $p = 0.01$; group x amplitude interaction, $F(5) = 2.3$, $p = 0.05$), this difference was modest (35°/sec, on average) and insufficient to explain the manifest hypometria.

If saccades are terminated prematurely, it is possible that peak velocity is not early enough in the saccade to observe differences between the two groups. Rather, the strong prediction of the early termination hypothesis is that velocities early in a saccade should be larger than expected for its amplitude. To test this, we examined the velocity of the eye 10 ms after saccade onset. We observed a trend inconsistent with the early termination hypothesis: velocity early in the saccade was similar to, or slightly less than, that of controls (Figure 6.3D, main

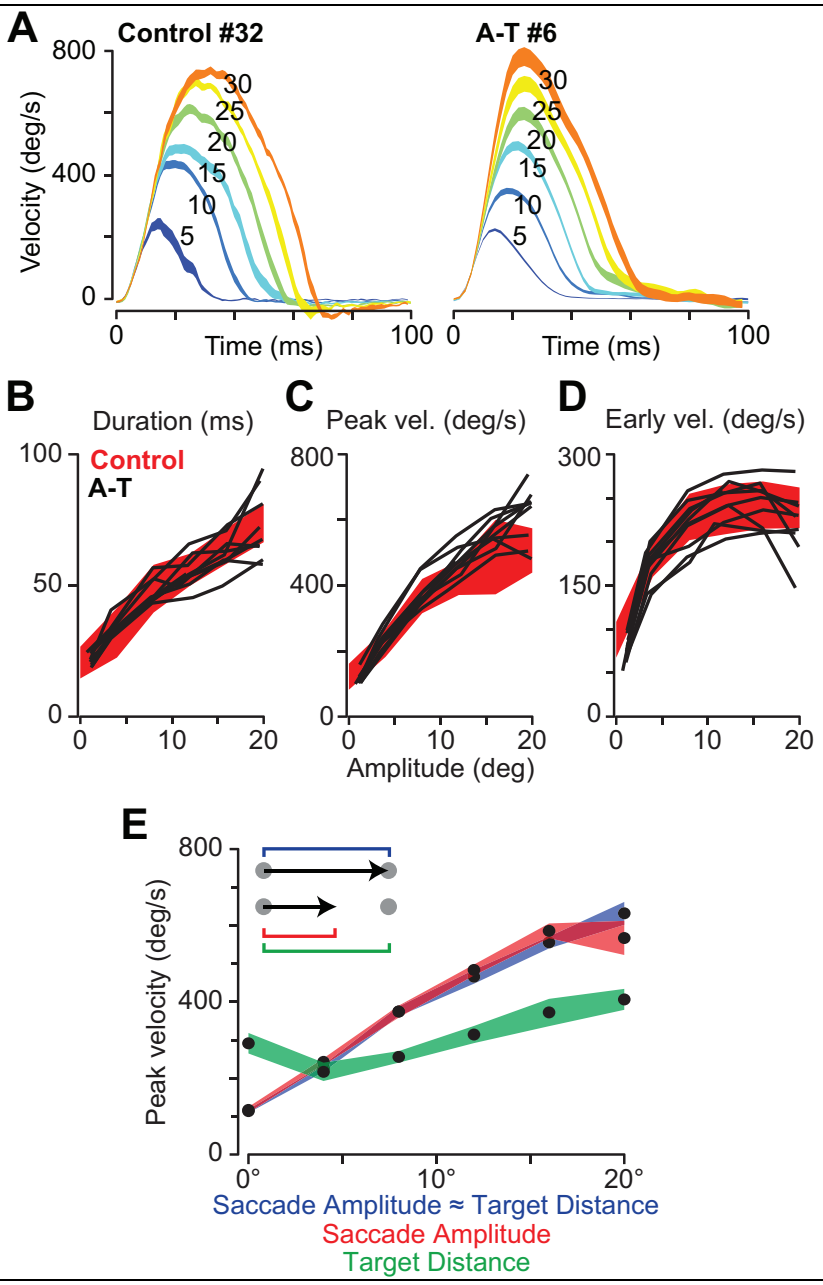

Figure 6.3: Saccade dynamics of A-T subjects did not indicate midflight termination of the movement. **A**, Velocity profiles for saccades of various amplitudes for representative control and A-T subjects with similar saccade peak velocities. Saccades were averaged within 5° bins, centered about 5° to 30°. Data are mean±SEM. The relationship between the saccade amplitude and **B**, Saccade duration, **C,** Peak velocity of saccades, and **D,** Early saccade velocity, 10 ms after onset. Control subjects are red (mean±1SD), and each black line is an A-T subject. **E**, Saccades of A-T subjects are appropriate for amplitude, not target distance. The relationship between amplitude and peak velocity is plotted for saccades that fully moved the eyes to the target (blue). The same relationship is plotted for saccades that did not move the eyes to the target, grouping them by saccade amplitude (red) or by target distance (green). Data are mean±SEM for A-T subjects.

effect of group: $F(1) = 1.3$, $p = 0.3$; group x amplitude interaction, $F(5) = 4.4$, $p < 0.01$). Thus, in summary, the duration, peak velocity, and early velocity of saccades in A-T subjects were inconsistent with the hypothesis that hypometria of saccades is a result of early termination.

Finally, to directly compare whether A-T subjects made saccades that intended to go to the target, or instead saccades that were appropriate for their actual amplitude, we conducted a within-subject analysis of the saccade dynamics of A-T subjects. First, we considered saccades that went to the target (Figure 6.3E, blue). While these saccades are uncommon, there is no ambiguity between the saccade amplitude and the distance to the target. These saccades were used as our reference. We next considered saccades that did not reach the target. When we grouped these movements by the target distance (Figure 6.3E, green), we observed a very different relationship between target distance and peak velocity than was apparent in the reference. If, on the other hand, these same movements were grouped by their amplitude (Figure 6.3E, red), we saw that their dynamics were indistinguishable from the reference saccades. That is, saccades across the experiment obeyed similar properties if grouped by their amplitude but exhibited different properties if grouped by the target distance. Conducting the same analysis using early saccade velocity similarly demonstrated that saccades were consistent in their dynamics for their actual amplitude, not the distance to the target (Figure 6.4). The saccade dynamics of

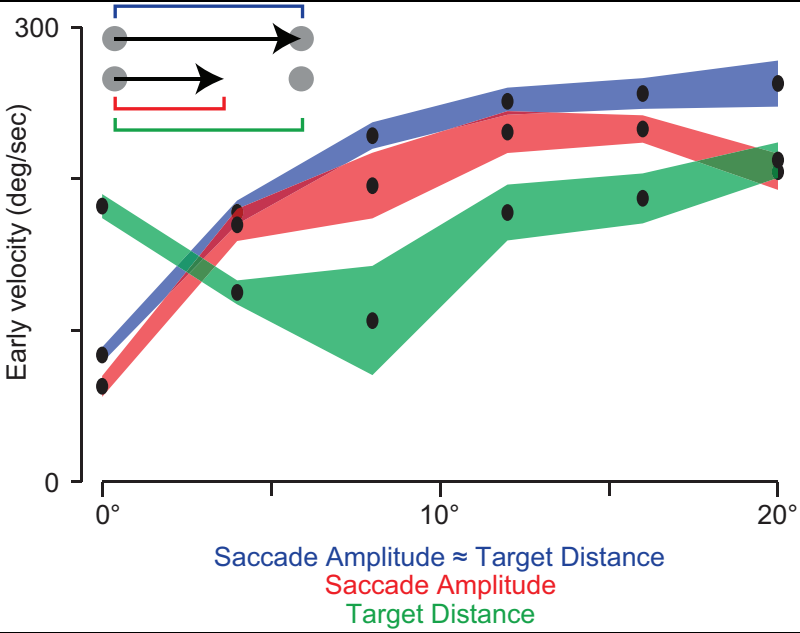

Figure 6.4: Early saccade velocity of A-T subjects was appropriate for amplitude, not target distance. The relationship between amplitude and velocity early in the saccade, 10 ms after onset, is plotted for saccades which moved to the target (blue). The same relationship is plotted for saccades which did not move to the target, grouping them by saccade amplitude (red) or by target distance (green). Data are mean±SEM for A-T subjects.

individuals with A-T were appropriate for their actual amplitude, appearing to be purposefully programmed to arrive short of the target. Why?

Signal dependent noise

We wondered if A-T subjects made a series of hypometric saccades as a strategy to minimize the deleterious effect of increased noise within their oculomotor system. We hypothesized that if A-T subjects exhibited greater signal dependent noise, employing a series of saccades instead of a single large saccade would be an effective compensatory response.

Signal dependent noise refers to the relationship between the standard deviation (SD) of the amplitude of an action, as a function of the mean amplitude

of that action. Intuitively, consider a situation where one must move the eye 40 degrees. A signal dependent noise of 0.5 indicates that a 20° saccade will be made with SD of 10°, while a 40° saccade will be made with a SD of 20°. Given this relationship, one has a choice: one could make a single saccade of 40±20° (mean±SD), and potentially be very inaccurate. Alternatively, one could choose to make a saccade of 20±10°, see what the result is, and then on average make a second saccade of 20±10°. Notably, if the first saccade results in a 30° movement, one will only move 10±5° subsequently – the second movement is selected after the consequences of the first are detected. This policy of making two saccades results in a net 40±10° amplitude – i.e. twice the accuracy of producing a single movement. In the presence of this form of noise, accuracy is improved by making several, smaller actions. One can continue to improve the net accuracy by subdividing the movement into 3, 4, or more actions. Therefore, in the presence of signal dependent noise, making a series of saccades is beneficial because it can reduce the endpoint variance.

We measured the noise of the saccades made by control and A-T subjects. Data from three example subjects are shown in Figure 6.5A, and data from the two groups are shown in Figure 6.5B. When a control subject made a small saccade, the standard deviation (SD) of the endpoint of that saccade was small; but when they made a larger saccade, the SD in the endpoint of that saccade increased commensurately. That is, noise was signal dependent in our subjects.

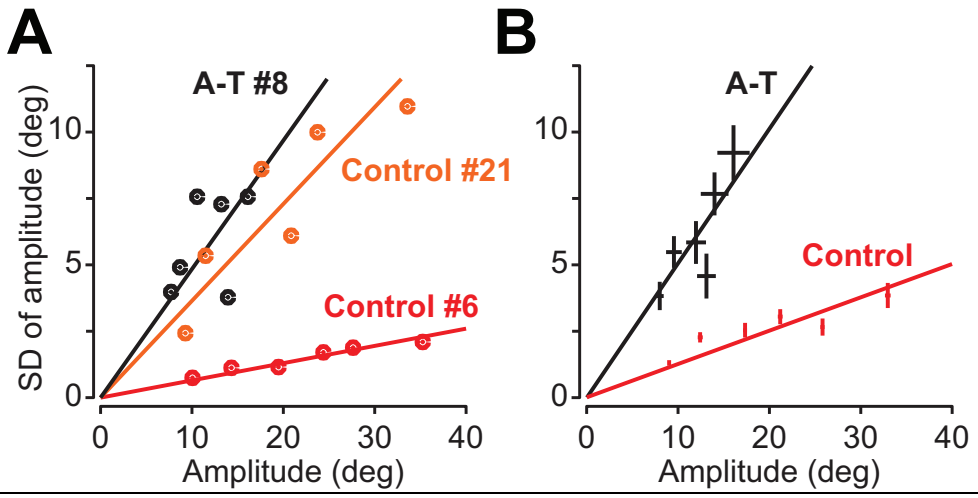

Figure 6.5: Signal dependent noise. **A,** The relationship between the mean amplitude of the first saccade made to each target and the standard deviation of that amplitude for two control and one A-T subject (same subjects as Figure 6.1B & C). Control #6 had a slope of 0.06, Control #2 had a slope of 0.36, and the A-T #8 had a slope of 0.48. **B,** Group data of the same relationship. The slope of the best fit line through the origin to each group is greater in A-T subjects (A-T: 0.51, Control: 0.13). Data in B are mean±SEM across subjects.

We quantified the relationship between mean saccade amplitude and the SD of the saccade amplitude. Both control and A-T subjects displayed an increase in the SD as a function of amplitude. However, A-T subjects exhibited a 4-fold greater SD for actions with the same mean amplitude. Control subjects had a signal dependent noise of 0.15±0.02 (fit to each subject, mean±SEM) while A-T subjects had a noise of 0.55±0.03 ($t(53) = 11.6$, $p = 4E\text{-}16$; alternatively, fit to each group's data, Figure 6.5, control noise: 0.13, A-T noise: 0.51, $t(9) = 11.6$, $p < 1E\text{-}6$). That is, A-T subjects had a significantly greater signal dependent noise than did control subjects.

If our only concern is to minimize endpoint variability, the best action is one that produces a long sequence of infinitely small saccades. This was not what we observed. However, in general the motor system also considers time as

a factor, seeking to complete movements quickly (Shadmehr et al., 2010). Making a sequence of small actions will improve accuracy but introduce undesirable delay in visualization of the target. One should select a policy – the number and amplitude of saccades – that will achieve the target both accurately and quickly.

Optimal Feedback Controller

We formalized our hypothesis by constructing an optimal feedback controller to identify the best movement policy to achieve the target. The controller included signal dependent noise, and was used to compute the policy that minimized a cost function. We imposed a quadratic accuracy cost (Harris and Wolpert, 1998), a hyperbolic time cost (Shadmehr et al., 2010; Shadmehr and Mussa-Ivaldi, 2012), and a quadratic effort cost (Todorov and Jordan, 2002). In general, missing the target, taking longer, and making larger saccades were penalized by each of these costs, respectively. Because visual feedback was available to subjects, the controller was fully observable; i.e. the consequences of a movement were assumed to be known after that movement was concluded – position is thus incorporated into the planning of the next movement. Using our model, for each subject we used their measured noise to predict the optimal number and gain of saccades for each target.

Consider three hypothetical subjects, one with low signal-dependent noise (Figure 6.6A and B, left), one with medium signal-dependent noise (Figure 6.6A and B, middle), and one with high signal-dependent noise (Figure 6.6A and B, right). The costs in Figure 6.6A depict the relative merits for various ways to arrive at a 10° target. At any level of noise, the end point accuracy cost is highest if the target is achieved with a single saccade, and declines if the target is achieved with a series of two or more saccades. In contrast, the time cost is lowest for a single saccade and grows larger with increasing number of saccades. As a consequence, the optimal number of saccades is one that minimizes the total cost and balances the competing demands of accuracy and time.

The model made three predictions. 1) For the subject with high signal-dependent noise, the accuracy cost is higher than it is for the subject with the low noise. As a consequence, the optimal number of saccades for the subject with high noise is three saccades, whereas the optimal number for the subject with low noise is a single saccade. In general, the model predicts that subjects with higher noise should make more saccades to a target of the same amplitude, accepting a delay in fixating the goal, i.e. a greater time cost, in order to improve their accuracy. 2) Subjects with higher noise are predicted to have a reduced saccade gain. For example, for a 10° target, the amplitude of the first saccade isexpected to be 7.6° for the high noise subjects and 9.8° for low noise subjects (Figure 6.6B). 3) The series of saccades is predicted to converge on the target,

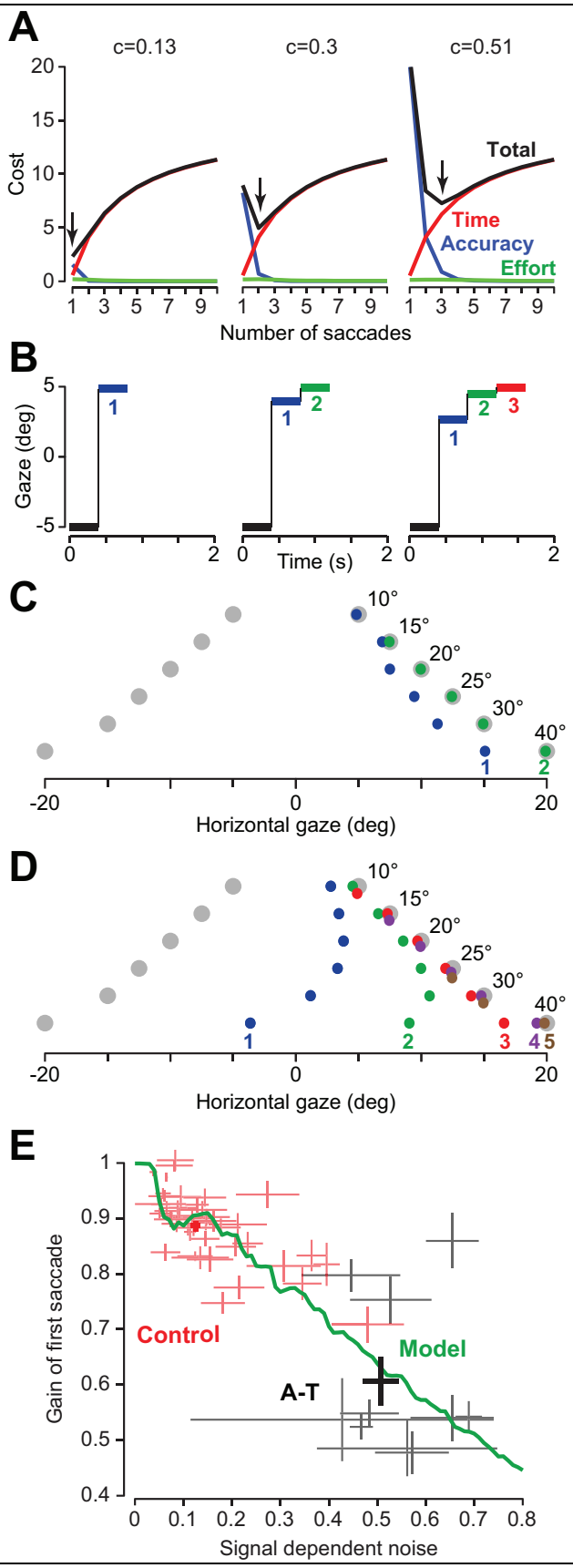

Figure 6.6: Predictions of an optimal feedback controller. **A,** An optimal feedback controller with accuracy (blue), time (red) and effort (green) costs predicts that an average control subject (c = 0.13) will make 1 saccade to a 10° target. Subjects with greater signal dependent noise (c = 0.3) are predicted to make 2 saccades while A-T subjects (c = 0.51) on average are predicted to make 3 saccades. **B**, The model predicts that subjects with increased noise (middle), including A-T subjects (right), will exhibit a reduced saccade gain and a converging series of saccades to the 10° target, as compared to control subjects with low noise (left). **C,** The predicted series of saccades for each target for the Control group (c = 0.13). **D,** The predicted series of saccades for each target for the A-T group (c = 0.51). The endpoints of each saccade are colored as in B. Targets (gray) are to scale. **E**, Predicted and observed initial saccade gain. Subjects (A-T, red; Control, black) with greater signal dependent noise have a smaller first saccade gain, as predicted by the model (green). Thin lines are individual subject noise and mean gain, error bars are standard error. Thick lines are group mean±SEM

with saccades of progressively decreasing amplitude for all targets (Figure 6.6C and Figure 6.6D). Across all targets, control subjects were predicted to exhibit an average gain of their first saccade of 87.7±0.9% (observed gain: 88.6±0.9%), while A-T subjects were predicted to exhibit an average gain of 60.9±2.1% (observed gain: 60.7±4.4%).

Importantly, the model not only made predictions regarding group differences, but it also made predictions about the saccade gains of individual subjects. For example, the model predicted that, even among the control subjects, those who exhibited greater signal dependent noise should also make a series of saccades (Figure 6.6A, middle), each with a reduced gain (Figure 6.6B, middle) as compared to subjects with lower noise. Indeed, this relationship is exactly what was observed – the least variable control subject had a signal dependent noise of 0.04 and a gain of 98% (predicted, 99.6%); while the most variable control subject had a signal dependent noise constant of 0.48 and a gain of 71% (predicted, 66%). Across all subjects, the relationship between the noise constant and saccade gain was well predicted by the model (Figure 6.6E, $r = 0.80$, $p = 2E\text{-}13$).

These results demonstrate that the series of saccades made by A-T subjects, and indeed the hypometric saccades sometimes made by control subjects, were not themselves a deficit in the ability to make saccades. Rather, the number and gain of saccades were well predicted by the signal dependent

noise of the oculomotor system. Making a series of saccades is an optimal policy in the face of increased signal dependent noise. Therefore, A-T subjects exhibit a primary deficit of increased signal dependent noise. The series of hypometric saccades in A-T, and in some healthy subjects, is not the primary deficit, but rather a compensatory adaptation to increased noise in the control of saccades.

Saccade gain in other movement disorders

Could this hypothesis also account for behavior observed in people with other neurologic disorders? To answer this question, we analyzed the variance of amplitude, gain, or endpoint error of the primary saccade from the data available in a number of previous reports. For each population, we estimated the increase in signal dependent noise and then used the model to predict the reduction in saccadic gain of that population.

Winograd-Gurvich et al. (2003) reported the gain and variance of saccades made by individuals with Huntington disease. When asked to make voluntary saccades to targets 20° apart, these subjects exhibited a 10% reduction in saccadic gain and an approximately 2.7 fold increase in signal dependent noise compared to age matched controls. Our model predicted a 9% reduction in gain for this increase in the amount of noise.

Van der Geest et al. (2006) measured the saccades of individuals with Williams syndrome and age matched controls in a saccade adaptation task.

During the baseline period, they observed a significant negative relationship, across subjects, between the noise of the oculomotor system and the gain of the first saccade made to a 20° target, i.e. subjects with greater noise were observed to have reduced saccadic gain. Affected individuals exhibited an 8% reduction in saccadic gain and a 2.6 fold increase in the signal dependent noise. Our model predicted a 10% reduction in gain for this increase in the amount of noise.

Ventre et al. (1992) asked individuals manifesting mild hemi-Parkinson disease to make saccades to targets at 20°. When targets were predictable and paced slowly (< 0.5 Hz) so that patients had adequate planning and initiation time, these patients exhibited a 2.3 fold increase in the signal dependent noise and a 15% reduction in gain. Our model predicted a 13% reduction in gain.

However, our model does not completely agree with all of the available data. In particular, Kapoula et al (2010) found that individuals with Lewy-Body dementia exhibited a 21% reduction in saccadic gain to targets placed 10° apart, with a 2.1 fold increase in the signal dependent noise. Our model predicted only an 8% reduction in gain to targets this close together. In the context of our model, this discrepancy may result from differences in the cost of time in these individuals who have both cortical lesions and dementia. Alternatively, impairments in target localization, or saccade planning or execution could be involved.

Section 6.4: Discussion

When asked to make a saccade to a target, healthy adults produced hypometric saccades, particularly when making large amplitude movements. That is, it appeared that the brain chose to make a movement that missed a visual target, and instead relied on a second, corrective saccade to fixate it on the fovea. People with ataxia-telangiectasia exhibited a more pronounced hypometria, making a series of saccades to achieve the target. Indeed, this behavior is not limited to A-T. Infants (Salapatek et al., 1980), older adults (Sharpe and Zackon, 1987; Irving et al., 2006; Litvinova et al., 2011), and patients with spinocerebellar ataxia (Bour et al., 2008; Christova et al., 2008), late onset cerebellar ataxia (Federighi et al., 2011), ataxia with oculomotor apraxia type 2 (Panouillères et al., 2013), familial myoclonic tremor with epilepsy (Bour et al., 2008), contralateral frontal eye field lesions (Rivaud et al., 1994), Williams syndrome (van der Geest et al., 2006), Alzheimer disease (Fletcher and Sharpe, 1986), Lewy Body dementia (Mosimann et al., 2005; Kapoula et al., 2010), Parkinson disease (DeJong and Jones, 1971; Mosimann et al., 2005; Litvinova et al., 2011), and Huntington disease (Avanzini and Girotti, 1979; Lasker and Zee, 1997; Winograd-Gurvich et al., 2003) all have been shown to exhibit more pronounced saccadic hypometria, with reduced saccadic gain and increased endpoint error after an initial saccade to a target.

Why does the brain of healthy adults and of patients with A-T choose this behavior, instead of making a single eye movement? We found that in the presence of signal dependent noise, one can be more accurate by producing hypometric movements. Actions of reduced amplitude have a distribution of consequences with lower variability, such that a series of smaller actions can result in increased accuracy. We hypothesized that this effect underlies the tendency of subjects with A-T to make a series of small saccades to shift their gaze. We observed that A-T subjects made 2-6 saccades to a target, typically moving only approximately halfway to the target with their first movement, and tended to make more saccades when asked to make larger movements. Control subjects made 1-2 saccades to the target, typically moving directly to, or near to, the target. Both groups displayed an amplitude dependent increase in the variance of the consequences of their eye movements, but the increase was 4-fold larger in A-T subjects. Using an optimal feedback controller to formalize our hypothesis, we found that the increased number and reduced gain of saccades of patients with A-T were well predicted by the signal dependent noise. In fact, control subjects who manifest larger signal dependent noise also exhibited a reduced saccadic gain, although more subtle than that of subjects with A-T, in agreement with the predictions of our model. The behavior of both patients with A-T and control subjects was the best possible way to accommodate the noise properties of their specific oculomotor systems.

This effect was not limited to patients with A-T, as our analysis of previously published data also found that in a number of other neurological disorders (Huntington disease, Williams syndrome, and Parkinson disease), the changes in saccadic gain were consistent with the model predictions, given only the measured increased in saccadic signal dependent noise. Therefore, in these conditions the brain may also purposefully choose to fall short of the target, as this was the best choice that simultaneously maximized accuracy while minimizing the time to fixate the target.

We have emphasized the role of noise on the optimal number and gain of saccades, as patients with A-T exhibited a dramatic increase in their oculomotor noise. It is important to note that differences in the cost of time and valuation of the stimulus across people or in patient populations would also have an effect on saccade gain. People who discount reward more rapidly would be expected to have larger saccade gain and make fewer saccades, on average. We assumed all subjects had the same cost of time and stimulus valuation for simplicity, but differences between individuals' valuation of the task and rate of discounting reward in time (Chapter 4; Choi et al., 2014) likely explain, in part, some of the additional variability within both populations and potentially in the behavior of other patient groups.

Our results demonstrate that saccadic hypometria in A-T, a seeming impairment in the control of saccades, instead likely reflects a compensation for

another underlying deficit, increased signal-dependent noise. Other symptoms in neurological disease, including those that are characteristic of A-T, may also reflect compensation for underlying deficits. For example, people with some neurological disorders exhibit an increase in the latency to initiation of a saccade to a target. While this may be a primary deficit in the ability to initiate movements, these individuals may instead be able to reduce their noise with additional planning. If so, an increase in latency could reflect the benefits of additional preparation time instead of an inability to initiate movements. Similarly, it is interesting to note that increasing the number and reducing the gain of saccades is not the only compensation available for increased signal dependent noise. Our hypothesis made use of the fact that lower amplitude movements have lower endpoint variability, so reducing the gain of saccades results in greater accuracy. In addition, in theory, producing slower movements should also result in reduced endpoint variability (Harris and Wolpert, 1998). Subjects could slow their movements to improve accuracy. While A-T subjects did not produce slower saccades and instead chose to be hypometric, patients with spinocerebellar ataxia type 2 produce dramatically slowed saccades (Bürk et al., 1999; Federighi et al., 2011) with relatively normal gain (Federighi et al., 2011). This symptom is potentially a compensation for increased noise. Given our results, these alternative hypotheses bear consideration.

Overall, our results demonstrate that saccadic hypometria in healthy controls and patients with movement disorders may be a compensatory adaptation to signal dependent noise properties of the oculomotor system. The brain of patients with A-T as well as healthy controls appears to understand the dynamics of the oculomotor control system and plans actions within the constraints of these dynamics. Individuals with A-T as well as healthy controls select the ideal actions for the bodies they inhabit.

Chapter 7: Conclusion

Each movement we make – of our eyes, our bodies, and our limbs – is the result of a complex process of decision making. The brain must select each of our actions, weighing its internal desires and knowledge of the world. The study of movements allows us a window into these factors: at each moment, how does the brain decide what movement to make, and why?

In this thesis, our goal was to use careful experiments to examine some of the factors underlying the selection of actions. By perturbing movements, creating artificial redundancies, pitting systems of learning against each other, and studying differences in the way different people choose to move, we explored the way three factors – errors, reward, and time – are involved in action selection.

We began by studying one role of error in action selection. In Chapter 2 and Chapter 3, we found that the policy underlying the selection of movement changes when the brain detects a change in the distribution of errors in the task. People typically learned from errors in the consequences of their actions, but in the presence of artificial redundancy could use other policies. We found that people explored; reinstated previously learned actions; and deinstantiated components of a learned behavior.

In subsequent chapters, we used paradigms where we opposed two factors to study the relationships between them. In Chapter 4, we found a dominant role for error, over reward, in the selection of behavior. When the error-based and reward-based learning systems were pit against each other, learning from error dominated behavior. In Chapter 5, we used a paradigm where time and reward were opposed to examine interindividual differences in the cost of time: waiting longer resulted in a greater probability of success. In doing so, we found that the way people value time correlated with the speed of their eye movements, suggesting a common cost of time in decision making and movement control. In Chapter 6, we found that the tradeoff between error and time could explain the way healthy people and people with a neurological disorder made eye movements. People with ataxia-telangiectasia and control participants used this tradeoff to select the best eye movements to shift their gaze quickly and accurately given the dynamics of their oculomotor systems.

While we focused on a subset of the movements made by the motor system in controlled settings, our hope is that these results provide insight into the mechanisms governing action selection more generally. In the laboratory, errors and rewards can be clearly controlled and defined, movements can be constrained and measured, and time is limited. Outside the lab, while these constraints are looser, these factors likely play similar roles.

Our work supports the idea the neural substrates of the perception and evaluation of time, of errors, and of rewards are likely distinct. Prior work has demonstrated a role for the cerebellum and basal ganglia, in part, for evaluation of prediction error and reward that result from a movement. Future study examining in greater detail the neural substrates of these factors, and in particular where and how they are integrated in the brain, would provide valuable insight into the control of movement.

Finally, it is our sincere hope that these results can be used to enhance the care of patients with neurological disorders. Better understanding of the mechanisms involved in decision making and motor control may aid the early and accurate diagnosis of neurological disease, improve our understanding of the underlying deficits, and guide the development and testing of treatments. Our hope is that these and future results can contribute not only to our understanding of the human brain, but also can be applied to improve the lives of our patients.

Appendix A: Making multiple saccades to a target: Optimal feedback controller

Problem statement

Let $x^{(k)}$ be the position of the eye with respect to the target. Saccade *k* is produced by motor command $u^{(k)}$, which moves the eye to the position denoted by $x^{(k+1)}$, but is affected by signal dependent noise:

$$\begin{aligned} x^{(k+1)} &= x^{(k)} + u^{(k)}\left(1 + c\phi^{(k)}\right) \\ \phi^{(k)} &\sim N(0,1) \end{aligned} \qquad \text{Eq. A1}$$

In Eq. (A1), ϕ is a normally distributed random variable with mean 0 and variance 1, and c is a constant. The expected value and variance of the saccade amplitude produced by the motor command $u^{(k)}$ are:

$$\begin{aligned} E\left[x^{(k+1)} \middle| x^{(k)}, u^{(k)}\right] &= x^{(k)} + u^{(k)} \\ \mathrm{var}\left[x^{(k+1)} \middle| x^{(k)}, u^{(k)}\right] &= c^2\left(u^{(k)}\right)^2 \end{aligned} \qquad \text{Eq. A2}$$

Let us assume that the total cost of producing a series of saccades $\mathbf{u} = \left[u^{(0)} \quad u^{(1)} \quad \cdots \quad u^{(n-1)}\right]$ is the sum of the costs of accuracy, time, and effort (Shadmehr et al., 2010; Shadmehr and Mussa-Ivaldi, 2012):

$$
\begin{aligned}
J &= J_X + J_T + J_U \\
J_X &= \left(x^{(n)}\right)^2 \\
J_T &= \alpha\left(1-\left(1+\beta T\right)^{-1}\right) \\
J_U &= \rho\sum_{k=0}^{n-1}\left(u^{(k)}\right)^2
\end{aligned}
\qquad \text{Eq. A3}
$$

In Eq. (A3), T is the total duration of the series of saccades, α and β are constants that define the hyperbolic relationship between duration and the temporal cost of the movement, and ρ is a constant associated with the relationship between magnitude of the motor command and its cost. The duration T can be defined as a function of the motor commands $\mathbf{u}$ by noting that there is an approximately linear relationship between saccade amplitude and saccade duration (Collewijn et al., 1988):

$$
T = \left(n-1\right)s + \sum_{k=0}^{n-1}\left(a + bu^{(k)}\right) \qquad \text{Eq. A4}
$$

In Eq. (A4), *n* denotes the number of saccades and *s* indicates a fixed inter-saccadic interval of time.

If the target is located at distance $|x|$, the mathematical problem is to find the number of saccades *n* and the sequence of motor commands $\mathbf{u}$ that bring the eyes to the target in such a way that minimizes the total cost *J* in Eq. (A3). The optimum number of saccades and motor commands are denoted by variables n^* and $\mathbf{u}^*$, respectively. We used $\alpha = 15$, $\beta = 0.8\ \text{s}^{-1}$, $\rho = 0.002$ for all subjects. The

saccade amplitude-duration relationship ($a = 24$ ms, $b = 2.5$ ms/°), intersaccadic interval ($s = 400$ ms), and signal dependent noise constant for control ($c = 0.13$) and A-T subjects ($c = 0.51$) were estimated from the measured data (see Section 6.3).

Optimal Feedback Controller

In order to derive an analytical solution for the optimal number of saccades n^* and sequence of motor commands $\mathbf{u}^*$, it is useful to approximate the cost of time J_T as a quadratic function of motor commands $\mathbf{u}$. We did this using a second order Taylor series expansion. We computed the time cost for a reference series of saccades $\mathbf{u}_{ref}$, and then wrote the time cost around this set point for any other series of saccades $\mathbf{u}$:

$$J_T(\mathbf{u}) \approx J_T\big|_{\mathbf{u}_{ref}} + \left(\frac{dJ_T}{d\mathbf{u}}\bigg|_{\mathbf{u}_{ref}} \right)^T (\mathbf{u}-\mathbf{u}_{ref}) + \frac{1}{2}(\mathbf{u}-\mathbf{u}_{ref})^T \frac{d^2 J_T}{d\mathbf{u}^2}\bigg|_{\mathbf{u}_{ref}} (\mathbf{u}-\mathbf{u}_{ref}) \quad \text{Eq. A5}$$

Now consider the special case in which the series of saccades $\mathbf{u}$ differs from the reference series $\mathbf{u}_{ref}$ only on the *k*-th saccade, with all other components of the two series equal. In that case, Eq. (A5) can be written as:

$$J_T(\mathbf{u}) \approx J_T\big|_{\mathbf{u}_{ref}} + \frac{dJ_T}{du^{(k)}}\bigg|_{\mathbf{u}_{ref}} (u^{(k)} - u_{ref}^{(k)}) + \frac{1}{2}(u^{(k)} - u_{ref}^{(k)})^2 \frac{d^2 J_T}{du^{(k)} du^{(k)}}\bigg|_{\mathbf{u}_{ref}} \quad \text{Eq. A6}$$

We can rewrite the cost of time in terms of the sequence of motor commands $\mathbf{u}$:

$$T(\mathbf{u}) = (n-1)s + na + b\sum_{j=0}^{n-1} u^{(j)}$$
$$J_T(\mathbf{u}) = \alpha - \alpha\left(1 + \beta T(\mathbf{u})\right)^{-1} \quad \text{Eq. A7}$$

And the resulting derivatives:

$$\left.\frac{dJ_T}{du^{(k)}}\right|_{\mathbf{u}_{ref}} = \alpha\beta b\left(1 + \beta T(\mathbf{u}_{ref})\right)^{-2}$$
$$\left.\frac{d^2 J_T}{du^{(k)}du^{(k)}}\right|_{\mathbf{u}_{ref}} = -2\alpha\beta^2 b^2\left(1 + \beta T(\mathbf{u}_{ref})\right)^{-3} \quad \text{Eq. A8}$$

This allows us to approximate the cost of time for any command $u^{(k)}$ as follows:

$$J_T^{(k)} \approx d^{(k)} + \tau^{(k)}u^{(k)} + r^{(k)}\left(u^{(k)}\right)^2 \quad \text{Eq. A9}$$

Where the constants in Eq. (A9) are defined as:

$$d^{(k)} = \frac{J_T|_{\mathbf{u}_{ref}}}{n} - \left.\frac{dJ_T}{du^{(k)}}\right|_{\mathbf{u}_{ref}} u_{ref}^{(k)} + \frac{1}{2}\left.\frac{d^2 J_T}{du^{(k)}du^{(k)}}\right|_{\mathbf{u}_{ref}} \left(u_{ref}^{(k)}\right)^2$$
$$\tau^{(k)} = \left.\frac{dJ_T}{du^{(k)}}\right|_{\mathbf{u}_{ref}} - \left.\frac{d^2 J_T}{du^{(k)}du^{(k)}}\right|_{\mathbf{u}_{ref}} u_{ref}^{(k)} \quad \text{Eq. A10}$$
$$r^{(k)} = \frac{1}{2}\left.\frac{d^2 J_T}{du^{(k)}du^{(k)}}\right|_{\mathbf{u}_{ref}}$$

We can represent the total cost of making a sequence of commands as the sum of a cost for each command. The total cost for sequence of commands $\mathbf{u} = \begin{bmatrix} u^{(0)} & u^{(1)} & \cdots & u^{(n-1)} \end{bmatrix}$ is:

$$J = \sum_{k=0}^{n-1} J^{(k)} \qquad \text{Eq. A11}$$

Where the cost for each command, the cost per step, includes the contribution of that saccade to the accuracy, time, and effort costs. We have the following cost per step:

$$\begin{aligned} J^{(k)} &= q^{(k)}\left(x^{(k)}\right)^2 + d^{(k)} + \tau^{(k)}u^{(k)} + \left(r^{(k)} + \rho\right)\left(u^{(k)}\right)^2 \\ q^{(k)} &= 1 \text{ if } k = n, \text{ otherwise } q^{(k)} = 0 \end{aligned} \qquad \text{Eq. A12}$$

Let us begin at the final time step $k = n$. Let $\pi^*\left(x^{(k)}\right)$ indicate the optimal command on the k-th time step, if the eye is at the position $x^{(k)}$. At the final time step $k = n$, the best action that we can perform is $u^{(n)} = 0$. That is, the optimal policy is:

$$\pi^*\left(x^{(n)}\right) = 0 \qquad \text{Eq. A13}$$

The value of the states achieved under this policy is:

$$V_{\pi^*}\left(x^{(n)}\right) = \left(x^{(n)}\right)^2 \qquad \text{Eq. A14}$$

Next, let us find the optimal policy for time step $n-1$. We begin by computing the expected value of the value of the last state, as a function of the previous state and action:

$$\begin{aligned} E\left[V_{\pi^*}\left(x^{(n)}\right)\middle| x^{(n-1)}, u^{(n-1)}\right] &= E\left[\left(x^{(n)}\right)^2\middle| x^{(n-1)}, u^{(n-1)}\right] \\ &= E\left[x^{(n)}\middle| x^{(n-1)}, u^{(n-1)}\right]^2 + \text{var}\left[x^{(n)}\middle| x^{(n-1)}, u^{(n-1)}\right] \\ &= \left(x^{(n-1)} + u^{(n-1)}\right)^2 + c^2\left(u^{(n-1)}\right)^2 \end{aligned} \quad \text{Eq. A15}$$

The cost that we need to minimize at time step $n-1$ is:

$$\begin{aligned} V_\pi\left(x^{(n-1)}\right) = d^{(n-1)} + \tau^{(n-1)}u^{(n-1)} + \left(r^{(n-1)} + \rho\right)\left(u^{(n-1)}\right)^2 \\ + E\left[V_{\pi^*}\left(x^{(n)}\right)\middle| x^{(n-1)}, u^{(n-1)}\right] \end{aligned} \quad \text{Eq. A16}$$

This cost is minimized by the following policy:

$$\begin{aligned} \pi^*\left(x^{(n-1)}\right) &= -g^{(n-1)}\left(\frac{1}{2}\tau^{(n-1)} + x^{(n-1)}\right) \\ g^{(n-1)} &= \left(r^{(n-1)} + \rho + c^2 + 1\right)^{-1} \end{aligned} \quad \text{Eq. A17}$$

The value function under this optimal policy can be computed from Eq. (A16) and Eq. (A17):

$$V_{\pi^*}\left(x^{(n-1)}\right) = d^{(n-1)} - \frac{1}{4}g^{(n-1)}\left(\tau^{(n-1)}\right)^2 - g^{(n-1)}\tau^{(n-1)}x^{(n-1)} + \left(1 - g^{(n-1)}\right)\left(x^{(n-1)}\right)^2 \quad \text{Eq. A18}$$

We note that as in $V_{\pi^*}\left(x^{(n)}\right)$, the above value function is a quadratic function of x:

$$
\begin{aligned}
V_{\pi^*}\left(x^{(n-1)}\right) &= w_0^{(n-1)} + w_1^{(n-1)} x^{(n-1)} + w_2^{(n-1)} \left(x^{(n-1)}\right)^2 \\
w_0^{(n-1)} &= d^{(n-1)} - \frac{1}{4} g^{(n-1)} \left(\tau^{(n-1)}\right)^2 \\
w_1^{(n-1)} &= -g^{(n-1)} \tau^{(n-1)} \\
w_2^{(n-1)} &= 1 - g^{(n-1)}
\end{aligned}
\qquad \text{Eq. A19}
$$

For time step $n - 2$, we similarly begin by computing the expected value of the value of the (n-1)-th state, as a function of the previous state and action:

$$
\begin{aligned}
E\left[V_{\pi^*}\left(x^{(n-1)}\right) \middle| x^{(n-2)}, u^{(n-2)}\right] &= w_0^{(n-1)} + w_1^{(n-1)} \left(x^{(n-2)} + u^{(n-2)}\right) \\
&+ w_2^{(n-1)} \left(x^{(n-2)} + u^{(n-2)}\right)^2 + w_2^{(n-1)} c^2 \left(u^{(n-2)}\right)^2
\end{aligned}
\qquad \text{Eq. A20}
$$

The cost that we need to minimize at time step $n-2$ is:

$$
\begin{aligned}
V_{\pi}\left(x^{(n-2)}\right) &= d^{(n-2)} + \tau^{(n-2)} u^{(n-2)} + \left(r^{(n-2)} + \rho\right)\left(u^{(n-2)}\right)^2 \\
&+ E\left[V_{\pi^*}\left(x^{(n-1)}\right) \middle| x^{(n-2)}, u^{(n-2)}\right]
\end{aligned}
\qquad \text{Eq. A21}
$$

This cost is minimized by the following policy:

$$
\begin{aligned}
\pi^*\left(x^{(n-2)}\right) &= -g^{(n-2)} \left(\frac{1}{2} \tau^{(n-2)} + \frac{1}{2} w_1^{(n-1)} + w_2^{(n-1)} x^{(n-2)}\right) \\
g^{(n-2)} &= \left(r^{(n-2)} + \rho + w_2^{(n-1)} c^2 + w_2^{(n-1)}\right)^{-1}
\end{aligned}
\qquad \text{Eq. A22}
$$

The value function under this optimal policy is:

$$
\begin{aligned}
V_{\pi^*}\left(x^{(n-2)}\right) &= d^{(n-2)} - g^{(n-2)} \left(\frac{1}{4}\left(\tau^{(n-2)}\right)^2 + \frac{1}{2} \tau^{(n-2)} w_1^{(n-1)} + \frac{1}{4}\left(w_1^{(n-1)}\right)^2\right) \\
&+ w_0^{(n-1)} + \left(w_1^{(n-1)} - g^{(n-2)} \tau^{(n-2)} w_2^{(n-1)} - g^{(n-2)} w_1^{(n-1)} w_2^{(n-1)}\right) x^{(n-2)} \\
&+ \left(w_2^{(n-1)} - g^{(n-2)} \left(w_2^{(n-1)}\right)^2\right)\left(x^{(n-2)}\right)^2
\end{aligned}
\qquad \text{Eq. A23}
$$

The above function is quadratic in x and can be written as:

$$
\begin{aligned}
V_{\pi^*}\left(x^{(n-2)}\right) &= w_0^{(n-2)} + w_1^{(n-2)}x^{(n-2)} + w_2^{(n-2)}\left(x^{(n-2)}\right)^2 \\
w_0^{(n-2)} &= d^{(n-2)} + w_0^{(n-1)} - g^{(n-2)}\left(\frac{1}{4}\left(\tau^{(n-2)}\right)^2 + \frac{1}{2}\tau^{(n-2)}w_1^{(n-1)} + \frac{1}{4}\left(w_1^{(n-1)}\right)^2\right) \\
w_1^{(n-2)} &= w_1^{(n-1)} - g^{(n-2)}\tau^{(n-2)}w_2^{(n-1)} - g^{(n-2)}w_1^{(n-1)}w_2^{(n-1)} \\
w_2^{(n-2)} &= w_2^{(n-1)} - g^{(n-2)}\left(w_2^{(n-1)}\right)^2
\end{aligned}
\qquad \text{Eq. A24}
$$

This illustrates that with the optimal policy, the value function remains quadratic, and therefore we can repeat this process to compute the optimal policy from the last time step to the first time step.

In summary, given the choice of *n* saccades, the optimal motor commands are $\mathbf{u}^* = \left[\pi^*\left(x^{(0)}\right) \quad \cdots \quad \pi^*\left(x^{(n-1)}\right)\right]$. The total expected cost for this policy is the value function $V_{\pi^*}\left(x^{(0)}\right)$. To find the optimal number of saccades n^*, we searched among potential candidates from $n = 1$ to $n = 10$ and found the one that had the minimum total expected cost.

We iteratively refined the reference series of saccades to obtain the best approximation to the time cost. We first assumed *n* saccades of equal amplitude would be made for $u_{ref}^{(j)}$. That is, for all *j* saccades in the series, $u_{ref}^{(j)} = \frac{-x^{(0)}}{n}$. We used that reference series of saccades to approximate the time cost and compute the optimal policy. Using that policy, we refined our estimate of $u_{ref}^{(j)}$ in the next iteration, used that to approximate the time cost, and recomputed the policy. We

repeated this procedure until the policy converged – $u_{ref}^{(j)}$ did not change between iterations – which typically required only 2-3 iterations.

References

Avanzini G, Girotti F (1979) Oculomotor disorders in Huntington's chorea. J Neurol Neurosurg Psychiatry 42:581–589.

Bahill A, Clark M, Stark L (1975) The main sequence, a tool for studying human eye movements. Math Biosci 24:191–204.

Baldeo R, Henriques D (2013) Dual adaptation to opposing visuomotor rotations with similar hand movement trajectories. Exp Brain Res 227:231–241.

Boder E, Sedgwick R (1958) Ataxia-telangiectasia a familial syndrome of progressive cerebellar ataxia, oculocutaneous telangiectasia and frequent pulmonary infection. Pediatrics 21:526–554.

Bour LJ, van Rootselaar AF, Koelman JHTM, Tijssen MAJ (2008) Oculomotor abnormalities in myoclonic tremor: a comparison with spinocerebellar ataxia type 6. Brain 131:2295–2303.

Buckholtz JW, Treadway MT, Cowan RL, Woodward ND, Li R, Ansari MS, Baldwin RM, Schwartzman AN, Shelby ES, Smith CE, Kessler RM, Zald DH (2010) Dopaminergic network differences in human impulsivity. Science 329:532.

Bürk K, Fetter M, Abele M, Laccone F, Brice a, Dichgans J, Klockgether T (1999) Autosomal dominant cerebellar ataxia type I: oculomotor abnormalities in families with SCA1, SCA2, and SCA3. J Neurol 246:789–797.

Bütefisch CM, Khurana V, Kopylev L, Cohen LG (2004) Enhancing encoding of a motor memory in the primary motor cortex by cortical stimulation. J Neurophysiol 91:2110–2116.

Cheng S, Sabes PN (2006) Modeling sensorimotor learning with linear dynamical systems. Neural Comput 18:760–793.

Chen-Harris H, Joiner WM, Ethier V, Zee DS, Shadmehr R (2008) Adaptive control of saccades via internal feedback. J Neurosci 28:2804–2813.

Choi JE-S, Vaswani PA, Shadmehr R (2014) Vigor of movements and the cost of time in decision making. J Neurosci 34:1212–1223.

Christova P, Anderson JH, Gomez CM (2008) Impaired eye movements in presymptomatic spinocerebellar ataxia type 6. Arch Neurol 65:530–536.

Collewijn H, Erkelens C, Steinmain R (1988) Binocular co-ordination of human horizontal saccadic eye movements. J Physiol 404:157–182.

Criscimagna-Hemminger SE, Shadmehr R (2008) Consolidation patterns of human motor memory. J Neurosci 28:9610–9618.

DeJong J, Jones G (1971) Akinesia, hypokinesia, and bradykinesia in the oculomotor system of patients with Parkinson's disease. Exp Neurol 68:58–68.

Diedrichsen J, White O, Newman D, Lally N (2010) Use-dependent and error-based learning of motor behaviors. J Neurosci 30:5159–5166.

Doya K (1999) What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? Neural Netw 12:961–974.

Eagle DM, Bari A, Robbins TW (2008) The neuropsychopharmacology of action inhibition: cross-species translation of the stop-signal and go/no-go tasks. Psychopharmacology (Berl) 199:439–456.

Emken JL, Benitez R, Sideris A, Bobrow JE, Reinkensmeyer DJ (2007) Motor adaptation as a greedy optimization of error and effort. J Neurophysiol 97:3997–4006.

Eysenck S, Pearson P, Easting G, Allsopp J (1985) Age norms for impulsiveness, venturesomeness and empathy in adults. Pers Individ Dif 6:613–619.

Farr AK, Shalev B, Crawford TO, Lederman HM, Winkelstein J a, Repka MX (2002) Ocular manifestations of ataxia-telangiectasia. Am J Ophthalmol 134:891–896.

Federighi P, Cevenini G, Dotti MT, Rosini F, Pretegiani E, Federico A, Rufa A (2011) Differences in saccade dynamics between spinocerebellar ataxia 2 and late-onset cerebellar ataxias. Brain 134:879–891.

Fischer B, Ramsperger E (1984) Human express saccades: extremely short reaction times of goal directed eye movements. Exp Brain Res 57:191–195.

Fletcher WA, Sharpe JA (1986) Saccadic eye movement dysfunction in Alzheimer's disease. Ann Neurol 20:464–471.

Frank MJ, Seeberger LC, O'Reilly RC (2004) By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306:1940–1943.

Fuchs AF, Scudder CA, Kaneko CRS (1988) Discharge patterns and recruitment order of identified motoneurons and internuclear neurons in the monkey abducens nucleus. J Neurophysiol 60:1874–1895.

Galea JM, 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. Cereb Cortex 21:1761–1770.

Ganesh G, Haruno M, Kawato M, Burdet E (2010) Motor memory and local minimization of error and effort, not global optimization, determine motor behavior. J Neurophysiol 104:382–390.

Gold J, Shadlen M (2002) Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. Neuron 36:299–308.

Gupta R, Ashe J (2007) Lack of adaptation to random conflicting force fields of variable magnitude. J Neurophysiol 97:738–745.

Haith AM, Krakauer JW (2013) Model-Based and Model-Free Mechanisms of Human Motor Learning. In: Progress in Motor Control (Richardson MJ, Riley MA, Shockley K, eds), pp 1–21 Advances in Experimental Medicine and Biology. New York, NY: Springer New York.

Haith AM, Reppert TR, Shadmehr R (2012) Evidence for hyperbolic temporal discounting of reward in control of movements. J Neurosci 32:11727–11736.

Harris CM, Wolpert DM (1998) Signal-dependent noise determines motor planning. Nature 394:780–784.

Harris CM, Wolpert DM (2006) The main sequence of saccades optimizes speed-accuracy trade-off. Biol Cybern 95:21–29.

Hirashima M, Nozaki D (2012) Distinct motor plans form and retrieve distinct motor memories for physically identical movements. Curr Biol 22:432–436.

Howard IS, Wolpert DM, Franklin DW (2013) The effect of contextual cues on the encoding of motor memories. J Neurophysiol 109:2632–2644.

Hoyt D, Taylor C (1981) Gait and the energetics of locomotion in horses. Nature 292:239–240.

Huang VS, Haith A, Mazzoni P, Krakauer JW (2011) Rethinking motor learning and savings in adaptation paradigms: model-free memory for successful actions combines with internal models. Neuron 70:787–801.

Huang VS, Shadmehr R (2009) Persistence of motor memories reflects statistics of the learning event. J Neurophysiol 102:931–940.

Hwang EJ, Smith MA, Shadmehr R (2006) Dissociable effects of the implicit and explicit memory systems on learning control of reaching. Exp Brain Res 173:425–437.

Hwang J, Kim S, Lee D (2009) Temporal discounting and inter-temporal choice in rhesus monkeys. Front Behav Neurosci 3:9.

Hyten C, Madden G, Field D (1994) Exchange delays and impulsive choice in adult humans. J Exp Anal Behav 62:225–233.

Irving EL, Steinbach MJ, Lillakas L, Babu RJ, Hutchings N (2006) Horizontal saccade dynamics across the human life span. Invest Ophthalmol Vis Sci 47:2478–2484.

Ivry RB, Hazeltine RE (1995) Perception and production of temporal intervals across a range of durations: evidence for a common timing mechanism. J Exp Psychol Hum Percept Perform 21:3–18.

Izawa J, Criscimagna-Hemminger SE, Shadmehr R (2012) Cerebellar contributions to reach adaptation and learning sensory consequences of action. J Neurosci 32:4230–4239.

Izawa J, Shadmehr R (2011) Learning from Sensory and Reward Prediction Errors during Motor Adaptation. PLoS Comput Biol 7:1–11.

Jimura K, Chushak MS, Braver TS (2013) Impulsivity and self-control during intertemporal decision making linked to the neural dynamics of reward value representation. J Neurosci 33:344–357.

Jimura K, Myerson J, Hilgard J, Braver TS, Green L (2009) Are people really more patient than other animals? Evidence from human discounting of real liquid rewards. Psychon Bull Rev 16:1071–1075.

Joiner WM, Smith MA (2008) Long-term retention explained by a model of short-term learning in the adaptive control of reaching. J Neurophysiol 100:2948–2955.

Kable JW, Glimcher PW (2007) The neural correlates of subjective value during intertemporal choice. Nat Neurosci 10:1625–1633.

Kagerer FA, Contreras-Vidal JL, Stelmach GE (1997) Adaptation to gradual as compared with sudden visuo-motor distortions. Exp Brain Res 115:557–561.

Kapoula Z, Yang Q, Vernet M, Dieudonné B, Greffard S, Verny M (2010) Spread deficits in initiation, speed and accuracy of horizontal and vertical automatic saccades in dementia with lewy bodies. Front Neurol 1:138.

Karniel A, Mussa-Ivaldi F (2002) Does the motor control system use multiple models and context switching to cope with a variable environment? Exp Brain Res:520–524.

Kasuga S, Hirashima M, Nozaki D (2013) Simultaneous processing of information on multiple errors in visuomotor learning. PLoS One 8:e72741.

Kawagoe R, Takikawa Y, Hikosaka O (1998) Expectation of reward modulates cognitive signals in the basal ganglia. Nat Neurosci 1:411–416.

Kim HF, Hikosaka O (2013) Distinct Basal Ganglia Circuits Controlling Behaviors Guided by Flexible and Stable Values. Neuron 79:1–10.

Kobayashi S, Schultz W (2008) Influence of reward delays on responses of dopamine neurons. J Neurosci 28:7837–7846.

Kording KP, Tenenbaum JB, Shadmehr R (2007) The dynamics of memory as a consequence of optimal adaptation to a changing body. Nat Neurosci 10:779–786.

Körding KP, Wolpert DM (2004) The loss function of sensorimotor learning. Proc Natl Acad Sci U S A 101:9839–9842.

Krakauer JW, Ghez C, Ghilardi MF (2005) Adaptation to visuomotor transformations: consolidation, interference, and forgetting. J Neurosci 25:473–478.

Krakauer JW, Ghilardi MF, Ghez C (1999) Independent learning of internal models for kinematic and dynamic control of reaching. Nat Neurosci 2:1026–1031.

Krakauer JW, Mazzoni P (2011) Human sensorimotor learning: adaptation, skill, and beyond. Curr Opin Neurobiol 21:636–644.

Krakauer JW, Mazzoni P, Ghazizadeh A, Ravindran R, Shadmehr R (2006) Generalization of motor learning depends on the history of prior action. PLoS Biol 4:e316.

Krakauer JW, Pine ZM, Ghilardi M, Ghez C (2000) Learning of Visuomotor Transformations for Vectorial Planning of Reaching Trajectories. J Neurophysiol 20:8916–8924.

Lasker AG, Zee DS (1997) Ocular motor abnormalities in Huntington's disease. Vision Res 37:3639–3645.

Lewis RF, Lederman HM, Crawford TO (1999) Ocular motor abnormalities in ataxia telangiectasia. Ann Neurol 46:287–295.

Litvinova AS, Ratmanova PO, Evina EI, Bogdanov RR, Kunitsyna AN, Napalkov DA (2011) Age-related changes in saccadic eye movements in healthy subjects and patients with Parkinson's disease. Hum Physiol 37:161–167.

Mazzoni P, Hristova A, Krakauer JW (2007) Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation. J Neurosci 27:7105–7116.

Mazzoni P, Krakauer JW (2006) An implicit plan overrides an explicit strategy during visuomotor adaptation. J Neurosci 26:3642–3645.

Miall RC, Jenkinson N, Kulkarni K (2004) Adaptation to rotated visual feedback: a re-examination of motor interference. Exp Brain Res 154:201–210.

Millar A, Navarick DJ (1984) Self-control and choice in humans: Effects of video game playing as a positive reinforcer. Learn Motiv 15:203–218.

Milstein DM, Dorris MC (2007) The influence of expected value on saccadic preparation. J Neurosci 27:4810–4818.

Mosimann UP, Müri RM, Burn DJ, Felblinger J, O'Brien JT, McKeith IG (2005) Saccadic eye movement changes in Parkinson's disease dementia and dementia with Lewy bodies. Brain 128:1267–1276.

Myerson J, Green L (1995) Discounting of delayed rewards: Models of individual choice. J Exp Anal Behav 64:263–276.

Navarick D (2004) Discounting of delayed reinforcers: measurement by questionnaires versus operant choice procedures. Psychol Rec 54:85–94.

Niv Y, Daw ND, Joel D, Dayan P (2007) Tonic dopamine: opportunity costs and the control of response vigor. Psychopharmacology (Berl) 191:507–520.

Opris I, Lebedev M, Nelson RJ (2011) Motor Planning under Unpredictable Reward: Modulations of Movement Vigor and Primate Striatum Activity. Front Neurosci 5:61.

Panouillères M, Frismand S, Sillan O, Urquizar C, Vighetto A, Pélisson D, Tilikete C (2013) Saccades and eye-head coordination in ataxia with oculomotor apraxia type 2. Cerebellum 12:557–567.

Patton JH, Stanford MS, Barratt ES (1995) Factor structure of the Barratt impulsiveness scale. J Clin Psychol 51:768–774.

Pekny SE, Criscimagna-Hemminger SE, Shadmehr R (2011) Protection and expression of human motor memories. J Neurosci 31:13829–13839.

Peng M, Ying S, Zee D, Barker P, Ying H (2012) Three-dimensional magnetic resonance analysis of extraocular muscles demonstrates gaze-specific contractile shape changes. In: Association for Research in Vision and Opthamology Meeting Abstracts.

Perlman S, Becker-Catania S, Gatti RA (2003) Ataxia-telangiectasia: diagnosis and treatment. Semin Pediatr Neurol 10:173–182.

Perlman SL, Boder E, Sedgewick RP, Gatti RA (2012) Ataxia-telangiectasia. In: Handbook of Clinical Neurology, 3rd ed. (Subramony SH, Dürr A, eds), pp 307–332. Elsevier.

Pine A, Seymour B, Roiser JP, Bossaerts P, Friston KJ, Curran HV, Dolan RJ (2009) Encoding of marginal utility across time in the human brain. J Neurosci 29:9575–9581.

Rabe K, Livne O, Gizewski ER, Aurich V, Beck a, Timmann D, Donchin O (2009) Adaptation to visuomotor rotation and force field perturbation is correlated to different brain areas in patients with cerebellar degeneration. J Neurophysiol 101:1961–1971.

Rivaud S, Müri RM, Gaymard B, Vermersch AI, Pierrot-Deseilligny C (1994) Eye movement disorders after frontal eye field lesions in humans. Exp Brain Res 102:110–120.

Salapatek P, Aslin RN, Simonson J, Pulos E (1980) Infant saccadic eye movements to visible and previously visible targets. Child Dev 51:1090–1094.

Schaefer S, Shelly I, Thoroughman K (2012) Beside the point: motor adaptation without feedback-based error correction in task-irrelevant conditions. J Neurophysiol 107:1247–1256.

Scheidt RA, Reinkensmeyer DJ, Conditt MA, Rymer WZ, Mussa-Ivaldi FA (2000) Persistence of motor adaptation during constrained, multi-joint, arm movements. J Neurophysiol 84:853–862.

Shabbott BA, Sainburg RL (2010) Learning a visuomotor rotation: simultaneous visual and proprioceptive information is crucial for visuomotor remapping. Exp Brain Res 203:75–87.

Shadmehr R, Brashers-Krug T (1997) Functional stages in the formation of human long-term motor memory. J Neurosci 17:409–419.

Shadmehr R, Mussa-Ivaldi F (1994) Adaptive representation of dynamics during learning of a motor task. J Neurosci 14:3208–3224.

Shadmehr R, Mussa-Ivaldi F (2012) Biological Learning and Control: How the Brain Builds Representations, Predicts Events, and Makes Decisions. MIT Press.

Shadmehr R, Orban de Xivry JJ, Xu-Wilson M, Shih T-Y (2010) Temporal discounting of reward and the cost of time in motor control. J Neurosci 30:10507–10516.

Sharpe JA, Zackon DH (1987) Senescent saccades: effects of aging on their accuracy, latency and velocity. Acta Otolaryngol 104:422–428.

Shmuelof L, Delnicki RJ, Huang VS, Haith AM, Mazzoni P, Krakauer JW (2011) Error-based adaptation is suppressed and model-free mechanisms activated when actions and outcomes are uncorrelated. In: Program No 83.12. Neuroscience 2011 Abstracts. Washington, DC: Society for Neuroscience.

Shmuelof L, Huang VS, Haith AM, Delnicki RJ, Mazzoni P, Krakauer JW (2012) Overcoming motor "forgetting" through reinforcement of learned actions. J Neurosci 32:14617–14621.

Smith M, Ghazizadeh A, Shadmehr R (2006) Interacting adaptive processes with different timescales underlie short-term motor learning. PLoS Biol 4:1035–1043.

Straube A, Fuchs A (1997) Characteristics of saccadic gain adaptation in rhesus macaques. J Neurophysiol 77:874–895.

Tachibana Y, Hikosaka O (2012) The Primate Ventral Pallidum Encodes Expected Reward Value and Regulates Motor Action. Neuron 76:826–837.

Takikawa Y, Kawagoe R, Itoh H, Nakahara H, Hikosaka O (2002) Modulation of saccadic eye movements by predicted reward outcome. Exp Brain Res 142:284–291.

Tanaka H, Sejnowski TJ, Krakauer JW (2009) Adaptation to visuomotor rotation through interaction between posterior parietal and motor cortical areas. J Neurophysiol 102:2921–2932.

Taylor JA, Ivry RB (2011) Flexible cognitive strategies during motor learning. PLoS Comput Biol 7:e1001096.

Taylor JA, Klemfuss NM, Ivry RB (2010) An explicit strategy prevails when the cerebellum fails to compute movement errors. Cerebellum 9:580–586.

Taylor JA, Krakauer JW, Ivry RB (2014) Explicit and Implicit Contributions to Learning in a Sensorimotor Adaptation Task. J Neurosci 34:3023–3032.

Thoroughman K, Shadmehr R (2000) Learning of action through adaptive combination of motor primitives. Nature 407:742–747.

Todorov E (2005) Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system. Neural Comput 17:1084–1108.

Todorov E, Jordan MI (2002) Optimal feedback control as a theory of motor coordination. Nat Neurosci 5:1226–1235.

Tseng Y-W, Diedrichsen J, Krakauer JW, Shadmehr R, Bastian AJ (2007) Sensory prediction errors drive cerebellum-dependent adaptation of reaching. J Neurophysiol 98:54–62.

Van der Geest JN, Haselen GCL, Frens MA (2006) Saccade adaptation in Williams-Beuren Syndrome. Invest Ophthalmol Vis Sci 47:1464–1468.

Vaswani PA, Shadmehr R (2013) Decay of motor memories in the absence of error. J Neurosci 33:7700–7709.

Ventre J, Zee DS, Papageorgiou H, Reich S (1992) Abnormalities of Predictive Saccades in Hemi-Parkinson's Disease. Brain 115:1147–1165.

Verbruggen F, Logan GD (2008) Response inhibition in the stop-signal paradigm. Trends Cogn Sci 12:418–424.

Verstynen T, Sabes PN (2011) How each movement changes the next: an experimental and theoretical study of fast adaptive priors in reaching. J Neurosci 31:10050–10059.

Wagner MJ, Smith M a (2008) Shared internal models for feedforward and feedback control. J Neurosci 28:10663–10673.

Westheimer G (1999) Discrimination of short time intervals by the human observer. Exp Brain Res 129:121–126.

Willis WT, Ganley KJ, Herman RM (2005) Fuel oxidation during human walking. Metabolism 54:793–799.

Winograd-Gurvich C., Georgiou-Karistianis N, Evans a, Millist L, Bradshaw J., Churchyard a, Chiu E, White O. (2003) Hypometric primary saccades and increased variability in visually-guided saccades in Huntington's disease. Neuropsychologia 41:1683–1692.

Wu HG, Miyamoto YR, Castro LNG, Olveczky BP, Smith M a (2014) Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. Nat Neurosci 17:312–321.

Xu-Wilson M, Zee DS, Shadmehr R (2009) The intrinsic value of visual information affects saccade velocities. Exp Brain Res 196:475–481.

Yang Y, Lisberger SG (2014) Role of plasticity at different sites across the time course of cerebellar motor learning. J Neurosci 34:7077–7090.

Zarahn E, Weston GD, Liang J, Mazzoni P, Krakauer JW (2008) Explaining savings for visuomotor adaptation: linear time-invariant state-space models are not sufficient. J Neurophysiol 100:2537–2548.

CURRICULUM VITAE

PAVAN VASWANI

Born on November 18, 1987 in Melbourne, FL

EDUCATION

Johns Hopkins University School of Medicine, 2009 – *present*
MD PhD Program, Department of Neuroscience

University of Washington, 2005 – 2009
B.S. Computer Science; B.S. Neurobiology; B.S. Biochemistry
with Distinction, *summa cum laude*

PEER-REVIEWED PUBLICATIONS

Vaswani PA, Crawford TO, Wright JM, Lederman HM, Shadmehr R. Dissociating motor symptoms as primary deficits vs. compensatory adaptation: Ataxia-telangiectasia (submitted)

Vaswani PA, Shmuelof L, Haith AM, Delnicki RJ, Huang VS, Mazzoni P, Shadmehr R, Krakauer JW. Exploratory escape from persistent residual errors in motor adaptation tasks. (submitted)

Herzfeld DJ, **Vaswani PA**, Marko MK, Shadmehr R (2014) A memory of errors in sensorimotor learning. Science. (in press)

Choi JES*, **Vaswani PA***, Shadmehr R (2014) Vigor of movements and the cost of time in decision making. The Journal of Neuroscience 34:1212–1223

Vaswani PA, Shadmehr R (2013) Decay of motor memories in the absence of error. The Journal of Neuroscience 33:7700–7709.

Conway H, Smith B, **Vaswani PA**, Matsuoka K, Rignot E, Claus P (2009) A low-frequency ice-penetrating radar system adapted for use from an airplane: test results from Bering and Malaspina Glaciers, Alaska, USA, Annals of Glaciology 50: 93–97.

Vaswani PA. Dissociating motor symptoms as primary deficits vs. compensatory adaptations: ataxia-telangiectasia. Pediatric Neurology Grand Rounds. Aug 2014.

Vaswani PA, Shadmehr R. The two learning systems in motor control: reward vs. sensory prediction errors. Society for Neuroscience 2014. (accepted)

Crawford TO, **Vaswani PA**, Wright JM, Lederman HM, Shadmehr R. Optimal control of saccades in patients with ataxia-telangiectasia. Society for Neuroscience 2014. (accepted)

Vaswani PA, Shadmehr R. Motor adaptation when maximizing reward is in conflict with minimizing error. Society for Neuroscience 2013

Herzfeld DJ, **Vaswani PA**, Shadmehr R. Sensitivity of motor adaptation depends on history of experienced errors. Translational and Computational Motor Control 2013.

Herzfeld DJ, **Vaswani PA**, Shadmehr R. Sensitivity of motor adaptation depends on history of experienced errors. Society for Neuroscience Nov 2013.

Herzfeld DJ, **Vaswani PA**, Marko M, Kojima Y, Soetedjo R, Shadmehr R. Sensitivity of Motor Adaptation Depends on History of Experienced Errors. Gordon Conference on the Cerebellum 2013.

Vaswani PA, Shadmehr R. Contextual change induces spontaneous recovery of motor memory. Society for Neuroscience 2012

Choi JE, **Vaswani PA**, Shadmehr R. The cost of waiting: Impulsivity and the vigor of saccades. Society for Neuroscience 2012.

Reppert T, **Vaswani PA**, Shadmehr R. Sensitivity of movement vigor to changes in rate of reward. Advances in Computational Motor Control 2012

Vaswani PA, Shadmehr R. Decay of Motor Memories. Sensorimotor Research Day 2012.

Vaswani PA, Matsuoka KM. Developing a system for radar profiling of glacial ice. Northwest Glaciology Conference 2005.

GRANTS

A-T Children's Project, "Analyzing eye and head movements in patients with ataxia-telangiectasia," Sept 2013

A-T Children's Project, "Assessing the dynamics of reaching movements in patients with ataxia-telangiectasia," Dec 2013

TEACHING EXPERIENCE

Guest lecture, "Optimal Control of Saccades in Ataxia-Telangiectasia", *Learning Theory*, Johns Hopkins University, Department of Biomedical Engineering, Spring 2014

Teaching Assistant, *Nervous System and Special Senses*, Johns Hopkins University School of Medicine, 2013

Teaching Assistant, *Brave New World: The scientific, economic, and social impact of Computer Science*, University of Washington, Department of Computer Science, Spring 2009