

**From the holdings of the Department of Special Collections , Princeton University Library,
Princeton, NJ**

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyright law. We do not charge any permission or usage fees for the publication of images of material in our collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff. For more information regarding copyright, credit, and citation please review the information on the Copyright, Credit and Citation Guidelines page of our website.

<https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines>. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. <https://library.princeton.edu/special-collections/ask-us>

Doctoral Dissertations, 2011-2026

Psychology

2018

Action Selection and Action Execution in Human Learning

Mcdougale, Samuel David

Princeton, NJ : Princeton University

<http://arks.princeton.edu/ark:/88435/dsp01gt54kq72q>

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. <https://www.copyright.gov/title17/title17.pdf>

Action Selection and Action Execution in Human Learning

Samuel David McDougale

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE BY

THE DEPARTMENT OF

PSYCHOLOGY

Advisers: Jordan A. Taylor, Yael Niv

June 2018

©Copyright by Samuel David McDougle, 2018. All rights reserved.

Abstract

Intelligent behavior requires knowing both what to do in a given situation, and how to do it. Knowledge of the requisite whats and hows in life is not acquired through a single process — often a variety of distinct neural systems guide an organism through the learning process. For illustration, consider a child first learning to tie her shoes: This relatively simple behavior requires a confluence of factors that each draw on different neural mechanisms, including, but not limited to, a declarative memory system that recalls the crucial instructions her mother gives her, a reinforcement system that rewards successful attempts and thus reinforces the corresponding maneuvers, and a calibration system that coordinates, maintains, and updates the necessary muscle commands to account for an ever-changing environment and body (e.g. a new pair of laces, cold fingers, etc.). The cooperation of these various systems is vital for the acquisition of skills.

The first stage of any non-reflexive motor behavior is defined by volition: What do I want to do? I'll refer to this stage as action selection. Investigations of selection processes will be presented in the context of simple, discrete value-based decisions (Chapter 2), and more complex sensorimotor decisions (Chapters 3 and 4). After selection of an action, the motor system must then attempt to successfully carry out that action. I'll refer to this process as action execution. Execution is characterized by the accuracy of the selected action, and represented in the neural activity that ultimately leads to repeatable spatio-temporal patterns of coordinated muscle activation. Execution will be discussed in the context of reaching movements. The overarching theme of this thesis is the interaction of selection and execution systems during human reinforcement and sensorimotor learning.

Acknowledgments

I would first like to thank my supervisor, Jordan Taylor, for his amazing attention, support, and scientific guidance. I couldn't have asked for a better adviser, mentor, and friend.

I'd also like to thank the other members of the Intelligent Performance and Action Lab at Princeton (past and present) for always providing a fun and stimulating research environment. In particular, I'd like to thank Krista Bond and Peter Butcher for their tireless help with data collection.

I also thank Yael Niv and Rich Ivry for supervising me and for all their priceless attention and advice, as well as Jordan, Yael, and Uri Hasson for their work on my thesis committee.

I thank Javier Medina, who supervised me as a lab technician many years ago, and inspired me to continue my career in science.

Thank you to Kelly for her love, patience, and humor, and to my parents for their unceasing support.

Relevant Publications by the Thesis Author

Several recent publications have been adapted into three sub-chapters of this thesis:

- 1) The main findings and methods presented in **Chapter 2.1** were published in 2016:

McDougle, S. D., Boggess, M. J., Crossley, M. J., Parvin, D., Ivry, R. B., and Taylor, J. A. (2016). Credit assignment in movement-dependent reinforcement learning. *Proceedings of the National Academy of Sciences*, 113 (24), 6797-6802.

- 2) The main findings and methods presented in **Chapter 3.1** were published in 2015:

McDougle, S. D., Bond, K. M., and Taylor, J. A. (2015). Explicit and implicit processes constitute the fast and slow processes of sensorimotor learning. *Journal of Neuroscience*, 35(26), 9568-9579.

and further discussed in a review in 2016:

McDougle, S. D., Ivry, R. B., and Taylor, J. A. (2016). Taking aim at the cognitive side of learning in sensorimotor adaptation tasks. *Trends in Cognitive Sciences*, 20(7), 535-544.

- 3) The main findings and methods presented in **Chapter 3.2** were published in 2017:

McDougle, S. D., Bond, K. M., and Taylor, J. A. (2017). Implications of plan-based generalization in sensorimotor adaptation. *Journal of Neurophysiology*, 118(1), 383-393.

Contents

1	Introduction	1
1.1	Learning contains multitudes	1
1.2	Decision-making and movement: what to do versus how to do it	3
2	Execution informs selection: credit assignment and reward prediction errors	7
2.1	Movement errors in value-based decision-making	8
2.1.1	Task design and method	10
2.1.2	Results and modeling	15
2.2	Multiplexed error signals within a single task context	23
2.2.1	Task design and method	23
2.2.2	Behavioral results	25
2.3	fMRI investigation of striatal prediction errors in the context of action execution	26
2.3.1	Task design and method	27
2.3.2	fMRI methods	31
2.3.3	Behavioral results	34
2.3.4	Modeling results	38
2.3.5	Imaging results	43
2.4	General discussion	47
3	Selection informs execution: controlled selection and implicit execution processes in motor learning	51
3.1	Measuring explicit and implicit motor learning	52
3.1.1	Task design and modeling methods	55
3.1.2	Behavioral and modeling results	60
3.2	Explicit action selection determines which movements are adapted	61

3.2.1	Task design and methods	63
3.2.2	Generalization results	65
3.2.3	An updated state-space model	69
3.3	General discussion	73
4	Action selection strategies: working memory and cognitive motor learning	76
4.1	The validity of reaction time data as a reflection of selection processes in sensorimotor adaptation	79
4.2	Evidence for working memory signatures in explicit selection processes . . .	81
4.3	Two dissociable working memory strategies in motor learning	84
4.3.1	Task design and method	85
4.3.2	Results	87
4.4	Limiting reaction time to reveal dynamics of action selection	91
4.4.1	Task design and method	92
4.4.2	Behavioral results	98
4.4.3	Modeling results	101
4.4.4	Revealing response caching in a constrained-RT experiment	108
4.5	Cerebellar and hippocampal roles in explicit motor learning	113
4.5.1	The cerebellum and mental rotation	114
4.5.2	The medial temporal lobe and flexibility in motor learning	120
4.6	General discussion	127
5	Conclusion	129
5.1	Proceduralization, knowledge, and cognitive contributions to skills	129
5.2	Causal models and their role in learning	132
5.3	Parallels between models of reinforcement learning and motor learning . . .	134
	References	138

1 Introduction

1.1 Learning contains multitudes

Over many years of work in cognitive psychology, different forms of learning have often been confined to particular algorithmic processes. For instance, operant learning was equated with the incremental reinforcement of rewarded actions (Thorndike, 1911), and motor learning was equated with the adaptation of internal models for motor control (Marr, 1969). With these narrow definitions, an observed behavior can be classified as the result of a single process, and studied as such. This has been proven to be an immensely fruitful way to study individual learning algorithms, and such distinctions can also be useful when focusing on specific questions relating to the implementation of those algorithms in specific neural circuits (Marr, 1969).

However, much research in the past several decades has revealed that specific forms of learning are often a tangled knot of multiple psychological and neural systems working in parallel. Learned behaviors are usually arrived at via a blend of different computations, and these computations can represent vastly different psychological processes.

Evidence for multiple learning systems has been provided by many fields of cognitive psychology and neuroscience, including the dissociation of model-based and model-free reinforcement learning (Daw et al., 2005; Gläscher et al., 2010), intentional and automatic forms of memory (Jacoby, 1991), rule-based versus information-integration category learning (Ashby and Maddox, 2005), and controlled versus automatic decision-making (Stroop, 1935). In fact, allusions to a multiple-process framework can be found in William James' (1890) meditations on habit versus "volitional deliberation."

In one example, category learning, Ashby and colleagues (Ashby et al., 1998) showed, through a series of careful psychophysical experiments, that when learning to categorize objects people appear to take two broad approaches: If the difference between two categories is

clearly delineated by a single observable dimension (e.g. vertical stripes on tigers vs. round spots on leopards), humans adopt a “rule-based” strategy, which tends to produce rapid learning and is explicitly reportable. However, when the difference between categories is defined by many dimensions (e.g. a tumor vs. a non-tumor on an X-ray), learning is mediated by an “information-integration” strategy that is slow, robust, and implicit. These processes can be mapped on to different neural systems (e.g. prefrontal-cortex mediated working memory for rule-based, and striatal reinforcement learning for information-integration; Ashby and Maddox, 2005). These results echo the idea of “multiple realizability” (Putnam, 1975; Fodor, 1975), which contends that an observed phenomenon (e.g. a person successfully categorizing objects) can arise from multiple underlying physical processes. This notion has been cited to emphasize the importance of careful studies of behavior for neuroscience (Krakauer et al., 2017).

The central purpose of this thesis is to characterize how action selection and action execution processes both contribute to learning, and explore their interaction. There are several important reasons to investigate this topic: First, it is crucial to revisit our assumptions about different forms of learning, and not overly constrain our models. Disparate fields are going through a reevaluation of extant oversimplifications – in the study of reinforcement learning, for example, recent work has made clear the distinction between the traditional concept of incremental trial-and-error learning versus controlled, prospective planning (Daw et al., 2005).

A similar revaluation is underway in the study of human motor learning. Motor learning is often relegated to the set of processes that implicitly modify motor execution over time. But the factors determining how we select actions in the first place, and the importance of context in that selection process, are crucial. For example, recent work in motor adaptation shows how explicit selection processes develop alongside implicit modifications of movement execution, and these processes appear to trade off based on the learning environment (Taylor

et al., 2014).

Updated models of learning are also important in the diagnosis and rehabilitation of psychological disorders. One recent study dissociated working memory-based learning from trial-and-error reinforcement learning, and showed how learning deficits in schizophrenia appear to be linked to the former but not the latter (Collins et al., 2014). In motor adaptation, a recent study showed how cerebellar atrophy affects both explicit and implicit motor learning components (Butcher et al., 2017). Such clinical insights are vital for designing treatment approaches.

There are also clear *in silico* applications for identifying and modeling distinct human learning processes. While modern AI techniques like deep learning have achieved massive success in recent years (Krizhevsky et al., 2012), critics contend that many key human capacities are not being captured by such models, including, for example, the ability to acquire abstract language-like rules given limited data (Marcus, 2018). Furthermore, data-hungry gradient-descent models have trouble generalizing what they learn to novel situations, which is a hallmark of human learning (Marcus, 2018). Understanding how multiple cognitive processes contribute to behavior throughout learning can inform the development of AI that goes beyond pattern recognition.

1.2 Decision-making and movement: what to do versus how to do it

The starting point for these investigations is the distinction between learning to choose better actions (selection) and learning to perform those actions better (execution).

The theoretical approach includes computational modeling techniques from reinforcement learning and motor learning, and the experimental approach includes psychophysics, neuropsychology, and non-invasive brain imaging (functional magnetic resonance imaging - fMRI).

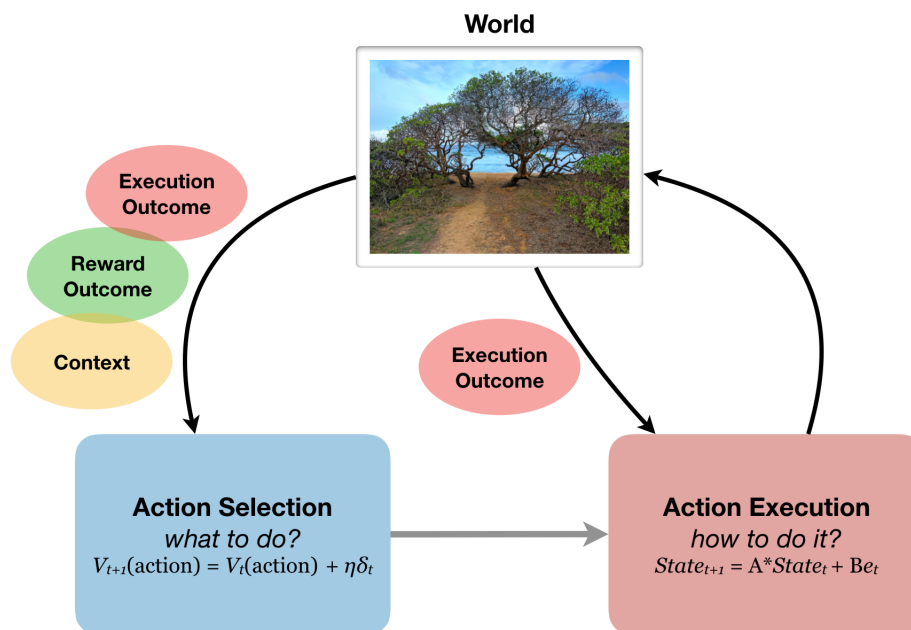


Figure 1.1: Overview

Action selection will be operationally defined as the adoption of a particular behavioral *policy* (Sutton and Barto, 1998). One policy could be, for example, to perform actions that currently have the maximum value relative to others. A model for learning this policy could be to track the values of different actions over time. After observing the outcome of an action (e.g. procuring a reward from the environment), its value (V) can be updated using a simple update rule (Figure 1.1) that learns from reward prediction errors - the difference between expected and observed reward. In this simple model, over time better actions will be selected among alternatives. This reinforcement learning process has been linked to phasic activity of dopaminergic neurons in the midbrain, which are thought to encode and broadcast a reward prediction error signal (Montague et al., 1996; Schultz et al., 1997). Using these signals, the brain can update the values of different actions to gradually maximize reward. This updating mechanism is thought to be implemented by dopamine-dependent plasticity in corticostriatal synapses (Barto, 1995; Reynolds et al., 2001). Using fMRI, dopamine reward prediction errors can be measured in humans, primarily in dopamine terminal areas

in striatal regions including the nucleus accumbens (NAc), and frontal regions including the ventromedial prefrontal cortex (vmPFC) and orbital frontal cortex (OFC; O'Doherty et al., 2003).

Once an action is selected, it must be executed. Execution can be defined by the particular transition from a desired goal to muscle activity. Like selection, learning how to execute a movement properly can also be modeled with a simple error-based process, where an internal model adapts to sensory prediction errors trial-by-trial. Sensory prediction errors represent the difference between expected and observed sensory outcomes. In this conception, the desired “state” is updated to reduce error over time (Figure 1.1).

The cerebellum is known to be vital for the execution of coordinated movements (Dalton, 1861; Thach et al., 1992), and has been signaled out as the primary locus of this kind of error-based motor learning process (Tseng et al., 2007; Schlerf et al., 2012). Learning in the cerebellum is thought to mainly rely on plasticity at the parallel-fiber to Purkinje cell synapse (Albus, 1971; Marr, 1969): Parallel fibers carry a state representation that includes sensory context and current outgoing motor commands (e.g., efference copy) to the cerebellar cortex. The second primary input to Purkinje cells, the climbing fibers, provide a teaching signal when an unexpected sensory event occurs (i.e., a sensory prediction error). Long term depression at the parallel fiber to Purkinje cell synapse induced by these teaching signals is thought to eventually lead to adapted motor behaviors that reduce the prediction error over time (Ito, 2006). In this simplified error-based motor learning model, an action will become progressively more accurate over time.

These two simple learning processes – reinforcement learning and error-based motor adaptation – form the jump-off point for many of the neural, behavioral, and theoretical investigations in this thesis. I will discuss how these processes may interact during learning, and also how additional processes beyond these simple iterative algorithms also play a role in multiple learning contexts. The additional processes discussed in this thesis will primarily

involve rapid and computationally demanding learning algorithms that appear to be fully separable from the slow, iterative processes highlighted above.

Chapter 2 concerns the effect of action execution on selection, in the context of a relatively simple decision-making situation – choosing between several lotteries. I ask how motor execution errors factor into the reinforcement learning process described above. Here, the selection policy is simplified to comparing the relative value of several stimuli. We find that motor errors are treated as a special case, and influence the decision making process in a unique way that does not appear to reflect a standard negative reward prediction error. This finding is observed in both behavioral and neural measures.

In **Chapter 3** I expand the concept of action selection to a common motor adaptation task, asking how selection processes proceed during motor learning, and how the selection system influences the execution system. In these tasks, the decision space is “parametric,” characterized by continuous directions of reaching movements. We find that selection and execution learning processes operate at different time scales, and appear to interact.

In **Chapter 4** I further investigate the nature of action selection processes in the context of motor learning, investigating neural substrates of motor selection strategies, and discussing how selection itself appears to have a deeper taxonomy consisting of multiple cognitive heuristics implemented in working memory. It appears that early in learning, action selection leverages a mental model of the learning environment. Later in learning, selection appears to follow a more simple “look-up table” strategy, where a model is no longer needed (perhaps reflecting a more “habitual” selection policy).

In **Chapter 5** I attempt to connect the results of Chapter 4 to the broader literature on instrumental learning, and provide one road map for future models of skill acquisition.

2 Execution informs selection: credit assignment and reward prediction errors

Research in value-based decision-making has emerged as an exceptional success in contemporary systems neuroscience (Schultz, 2013). The discovery and computational characterization of midbrain reward prediction errors (Schultz, 1998; Sutton and Barto, 1998) – which track the difference between expected and obtained rewards – has proved to be a resilient theoretical framework of instrumental learning in animals and *in silico*. The canonical form of this model describes how trial-and-error learning transpires in simple decision-making scenarios, where internal value representations of different choices are iteratively updated by reward prediction errors, leading to the selection of actions that yield the greatest average rewards.

Consider an agent attempting to glean reward (e.g. calories, money, etc.) from two stimuli, a and b . Across a series of time points, t , the agent samples, and updates the value, V , of each choice,

$$V_{t+1}(a) = V_t(a) + \eta\delta_t \quad (2.1)$$

$$\delta_t = r_t - V_t(a) \quad (2.2)$$

where $V_t(a)$ is the value of choice a at trial t , η is the learning rate, and δ is the reward prediction error, the difference between expected (V_t) and observed (r_t) reward.

This basic model assumes a deterministic map between action and reward outcomes. Imagine that actions a and b represent, respectively, reaching for and tasting an apple sitting in the left side of your fridge vs. a pear sitting in the right side of your fridge. Here, a and b are proxies for both the action (e.g. reaching to the left vs. reaching to the right) and the actual stimulus being selected (e.g. the apple vs. the pear). Thus, every trial outcome (each bout of reaching and tasting) provides information about the reward state of

the full stimulus/action pair. However, in the real world, the success of our actions is not a given. What if you reach for the apple and knock it off the shelf? Here, your decision itself may have been perfectly fine (you prefer the apple to the pear), but your execution of that decision was unsuccessful.

The following section poses a simple question: What happens to reward prediction error-driven learning when the movement that bridges the decision and the reward outcome is not successful? That is, where does action execution fit in to the standard reinforcement learning framework?

Study 2.1 uses a modified decision-making task that mixes reward and motor prediction errors. We find that motor errors appear to provide a salient contextual cue that allows the reinforcement learning (RL) system to separate external from internal sources of error. Several models are discussed to account for the interaction of motor errors and reward prediction errors. In study 2.2, the task design is streamlined and the predictions honed. In study 2.3, results from a model-based fMRI experiment are described, with a focus on prediction error processing in the striatum.

2.1 Movement errors in value-based decision-making

The premise for this study is as follows: Consider a diner at a coffee shop, reaching for her cappuccino. In a strict reinforcement learning scenario, if she reaches for her drink and accidentally spills it she would decrease the value of her cappuccino. This assumption is based on the fact that in simple reinforcement learning models, the underlying cause of the reward outcome is not specified: For the basic model, a decision was rendered (“I choose to sip my cappuccino”) but the expected reward was not received (“I have not ingested my caffeine”).

Of course, we’d expect our diner not to change how she values cappuccino based on an obvious movement error. Perhaps we could assume a salient motor error would “abort”

any update concerning the cappuccino. How can we can modify our model of reinforcement learning to account for this seemingly simple case of “credit assignment?” Does the reinforcement learning system “know” when action execution fails? Is this knowledge incorporated into the updating of value during action selection?

Humans are highly capable of tracking and updating the value of stimuli during decision making tasks, adjusting their choice behavior on the basis of reinforcement history (Daw et al., 2006). In laboratory studies of reinforcement learning, the underlying cause of reward outcomes is typically controlled, and explicitly described to participants as the result of probabilistic lotteries, with stimulus-specific reward probabilities that change over time. The stimuli used are often referred to as “bandits” or “slot machines.” Controlling how subjects conceive of such tasks is important for interpreting their decisions. In the standard version of this task, where participants make button presses to choose between different virtual slot machines, researchers can study how participants assess the value of different stimuli over time and adjust their choices accordingly. A rich body of work on human reinforcement learning and its neural implementation has resulted from these types of tasks (Daw et al., 2006; Gläscher et al., 2010; Niv et al., 2012; Wimmer et al., 2014; Bornstein et al., 2017; Leong, Radulescu, et al., 2017).

Here we set out to test the simple intuitive hypothesis that, during reinforcement learning, salient sensorimotor error signals could indicate when negative outcomes should be attributed to failures of the motor system; that is, a choice would not be devalued if it was clearly the *implementation* of that choice that was unsuccessful.

To test this, we developed a task similar to the slot machine task described above, but where outcomes could be assigned to either extrinsic properties of the environment (probabilistic rewards from the slot machines) or intrinsic motor error. We find that the presence of salient movement errors has a marked effect on choice behavior. This effect can be formalized by an attenuation of learning from negative reward prediction errors in the face

of motor errors. Specifically, the learning rate that scales the negative reward prediction error on motor error trials is comparatively smaller in conditions with salient motor errors versus conditions with no obvious motor errors, like the standard button-pressing task. Moreover, this effect appears to be weaker in individuals with cerebellar degeneration, consistent with a theoretical consideration that an internal estimate of motor competence may modulate reinforcement learning in certain contexts. (Note: the results described in this section have been published in McDougle et al., 2016.)

2.1.1 Task design and method

Participants performed a two-armed “bandit task” (Figure 2.1A), seeking to maximize points that were later exchanged for money. For all participants, the outcome of each trial was predetermined by two functions: One function defined if a target yielded a reward for that trial (“hit” or “miss”), and the other specified the magnitude of reward on hit trials (Figure 2.1B). The expected value was equivalent for the two targets on all trials; however, hit probability was not.

Our conditions were designed to provide different contexts for perceived errors. We hypothesized that in a discrete button-press condition (Standard condition) where motor errors are precluded, there would be no credit assignment problem and all reward outcomes would be attributed to stochasticity in the values of the choice stimuli. However, in the continuous reaching conditions (Spatial and Binary Conditions), where erroneous trials could be clearly attributed to motor errors, we reasoned that unrewarded outcomes on motor error trials would be attributed to the execution system, aborting any update of the value of the chosen stimulus. Thus, we predicted that changes in how participants inferred the “cause” of different choice outcomes would fundamentally change learning and decision-making.

Participants in all experiments provided informed consent, approved by the Institutional Review Boards at Princeton University and the University of California, Berkeley. In the

first experiment, 62 healthy adults were recruited from the research participation pool at Princeton University. All participants were right-hand-dominant according to the Edinburgh Handedness Inventory (Oldfield, 1971).

Participants were assigned to one of three groups ($n = 20$ participants per group). In the Standard group (Figure 2.1C), two red targets were positioned 8 cm from the center of a vertical screen. After a 1s delay, the word “Go” was displayed. Participants selected one of the two targets by pressing the left or right arrow on a keyboard with either the index or middle finger of the right hand. Participants were told that, on some trials, the selected target would turn green (“hit”) and display points earned on that trial (1 to 100). On other trials, the target would turn yellow (“miss”), earning them zero points. Instructions emphasized that participants had no control over whether a hit or miss occurred. A cumulative score was displayed in the center of the display. Participants were instructed to maximize their total points, which would later be exchanged for money (1-5 dollars, USD).

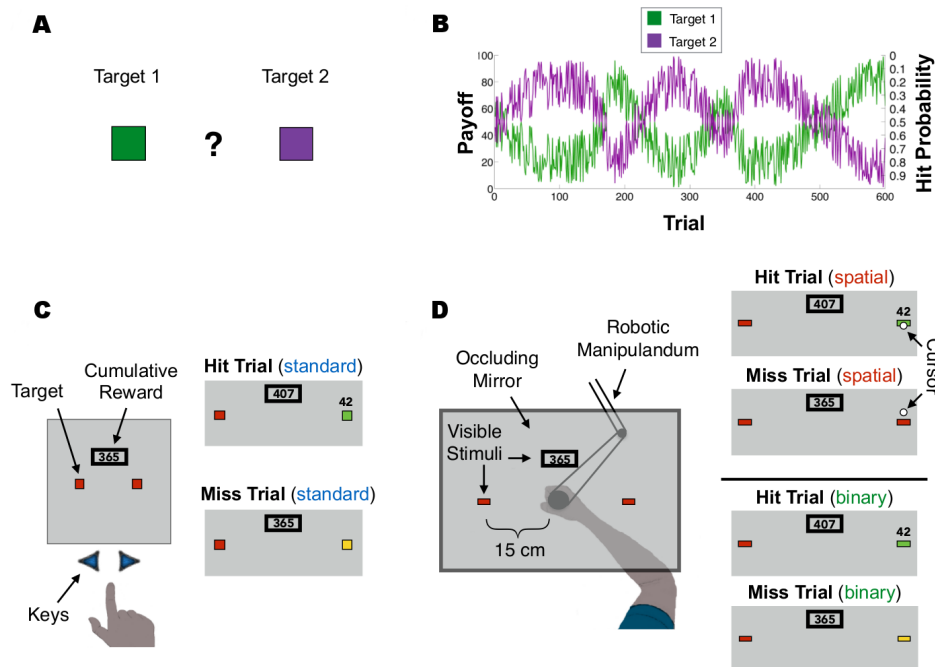


Figure 2.1: Task Design. (A) Participants performed a two-armed bandit task, choosing between two targets to maximize monetary payoff. (B) Two reflected, noisy sinusoids defined the payoff value (left axis) and probability of reward (hit, right axis, inverted) for the targets. (C) In the Standard condition, participants selected targets by pressing the left or right arrow keys on a keyboard. Example hit and miss trials are shown on the right. (D) In the Spatial and Binary conditions, participants reached to the selected target using a robotic manipulandum. Vision of the hand was occluded. In the Spatial condition, a small cursor appeared after the hand passed the target. On hit trials, the cursor overlapped with the target; on miss trials, the cursor appeared outside the target. Feedback in the Binary condition matched the Standard condition.

In the Spatial and Binary groups (Figure 2.1D), participants indicated their choices by using their right arm to move a robotic manipulandum (sampling rate = 1 kHz). At the start of each trial, the manipulandum moved the participant's hand to a starting point in the middle of the workspace, 35 cm from the participant's body. When the hand was within 5 cm of the starting position, participants received veridical cursor feedback. After maintaining the start position for at least 500 ms, the participant made a rapid reaching movement to the target of their choice.

The instructions for the two reach groups emphasized that the trial outcome (hit or miss) was dependent on reach accuracy: Points would be earned on trials in which they

hit the target and withheld on trials in which they missed. The stimuli were displayed on a horizontally mounted LCD screen reflected onto a mirror that was positioned above the movement surface. As such, vision of the hand was occluded. To make the motor control requirements demanding, targets were positioned 15 cm to the left and right of the start position and were limited in size (1 cm x 0.5 cm). In addition, the reach amplitude had to exceed 15 cm within 400 ms or a “Too Slow” message was displayed, and the trial was aborted and subsequently repeated.

In the Spatial group, a cursor (radius 0.5 cm) was displayed when the hand traversed the vertical axis of the target (Figure 2.1D). The position of the feedback corresponded to the position of the hand (subject to the constraints described below). Hits were defined as trials in which the feedback fell within the target region; misses were defined as trials in which the feedback fell outside the target region. The position of the cursor was veridical if the actual reach outcome matched the predetermined outcome. If the actual and predetermined outcome did not match, the position of the feedback cursor was subtly shifted to induce the predetermined outcome. This was accomplished by taking the location of the hand when it crossed the vertical axis of the target and adding a small translational shift, either away from the target for predetermined misses or toward the target for predetermined hits. The size of the shift was taken from a Gaussian distribution ($\mu = 0, \sigma = 0.5$ cm) relative to the location of the hand along the vertical axis. In this way, the perturbed feedback was correlated with the actual hand position. To emphasize the veracity of the feedback, trials were only valid if, during the reach, the hand stayed within an invisible 4 cm horizontal stripe centered about the target. If not, the screen displayed the message “Too Far,” and the trial was aborted and subsequently repeated.

The purpose of the Binary group was to investigate how the precision of motor error feedback affected decision-making. Specifically, we hypothesized that the less salient the motor error, the less clear the credit assignment problem would be. In the Binary condition,

participants also reached to indicate their choices but were not provided with cursor feedback; they received the same feedback as participants in the Standard condition (Figure 2.1D). Eliminating cursor feedback should yield a less salient motor error. All other aspects of visual and auditory feedback for the Spatial and Binary groups matched the Standard condition.

Point values and hit probabilities for each target were established by bounded pseudo-sinusoidal functions (Figure 2.1B). By mirroring both functions, the expected values of the two targets were matched on every trial. The outcomes (hit/miss, points) for each target on each trial were predetermined and identical for participants in all conditions.

We designed the reward functions so that, during alternating phases, one target was “high-miss high-reward (HMHR)” and the other target was “low-miss low-reward (LMLR).” Specifically, the HMHR target was associated with a greater probability of missing the target but a higher average magnitude of reward (points) on hit trials. The LMLR target was the opposite – there was a lower probability of missing the LMLR target, but lower average magnitude of rewards on hit trials. Which target was HMHR and which was LMLR varied over time: For example, for the first 165 trials, target 1 was HMHR, with higher miss probabilities (and higher payoffs), and target 2 was LMLR, with lower miss probabilities (and lower payoffs). The value/probability functions crossed paths at multiple points over the 600-trial block, creating reversals as to which target was the HMHR/LMLR choice. The locations of targets 1 and 2 were fixed for the entire session and counterbalanced across participants.

We also hypothesized that the cerebellum, a key region for accurate motor execution and error-based adaptation, would play a role in our task. Specifically, we hypothesized that individuals with damage to the cerebellum would behave similarly across the three conditions, either because, a) implicit processing of errors in the cerebellum (Tseng et al., 2007) plays a direct role in the proposed credit assignment process, making the relative salience of motor errors across conditions irrelevant, or b) individuals have lower expectations for their own

movement accuracy and thus believe they have less control over improving their reaching behavior in the reaching conditions, biasing them toward the LMLR choice.

In the patient experiment, 12 individuals with cerebellar degeneration (10 right-handed, two left-handed; mean age: 56, range 25-68) were recruited from the Princeton and Berkeley communities. Seven individuals had an identified genetic subtype, and five were of unknown etiology. Two patients diagnosed with spinocerebellar ataxia type 3 (SCA-3) were excluded from the final analysis given that phenotypes of SCA-3 may also show degeneration and/or dysfunction of the basal ganglia (Eichler et al., 2011), a region strongly implicated in reinforcement learning (Barto, 1995; Montague et al., 1996). Patients were screened using medical records and evaluated with the International Cooperative Ataxia Rating Scale and the Montreal Cognitive Assessments (MoCA). A control group of neurologically healthy adults, matched in terms of handedness and age, was also tested (mean age 64, range 50-72). Each participant performed all three conditions (400 trials each) of the first experiment (Figure 2.1) in a single session, using a graphical tablet (Wacom), and the order of tasks was counterbalanced across participants.

2.1.2 Results and modeling

Because both the HMHR and LMLR targets have equal expected value on every trial, observed biases toward one or the other could inform us about how participants treat different outcomes: A bias toward the current HMHR target in either reaching condition (Binary or Spatial) would suggest that motor errors (misses) are tolerated, because the HMHR target has a large number of misses but larger rewards on hit trials (Figure 2.1B). On the other hand, a bias toward the LMLR target in either reaching condition would suggest that motor errors cause a strong negative update. In the Standard condition, where motor errors are not considered, a bias toward the HMHR target would reflect high reward seeking, and a bias toward the LMLR target would reflect avoidance of non-rewarded trials.

Choice biases varied dramatically across conditions (Figure 2.2A; main effect: $p < 0.001$, all paired comparisons, $p < 0.01$). The Standard group exhibited a strong bias toward the LMLR target that was significantly different from the neutral bias value of 0.5 ($p < 0.001$), preferring choices with high hit probabilities and low reward values over larger rewards that are less frequent. In contrast, the Spatial group exhibited a markedly different pattern, preferring the HMHR target ($p < 0.001$). The opposing biases in the Standard and Spatial groups were consistent throughout the session, with participants tracking the respective targets (Figure 2.2B). The bias in the Binary condition was not significantly different from chance ($p = 0.59$). These results are consistent with the hypothesis that movement-related error feedback can signal when negative outcomes should be attributed to erroneous motor execution versus erroneous choice.

Critically, movement direction on trial $t + 1$ following miss trials was negatively correlated with the signed cursor error on trial t ($p < 0.001$), suggesting that, in terms of movement execution, participants were learning from their errors in the Spatial condition and adjusting their movements accordingly (Figure 2.2C).

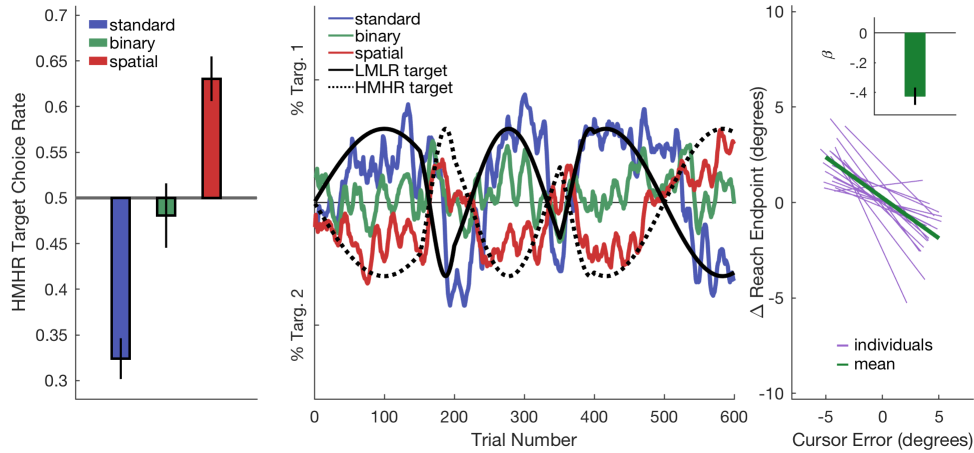


Figure 2.2: Behavioral results. (A) HMHR (“high-miss high-reward”) target choice rate was calculated by dividing the number of HMHR choices (choosing the target with the higher miss probability on that trial) by the total number of trials. (B) Group-averaged target preferences over time. Choices were backward-smoothed with a 10-trial bin. Black lines represent the current HMHR choice (dashed) and LMLR “low-miss low-reward” choice (solid). (C) Mean change in reach direction, relative to error direction, calculated on a trial-by-trial basis in the Spatial condition. The data here are restricted to pairs in which the first trial was a miss trial due to false feedback. The mean function (green) is a line given by the average of the group’s regression coefficients, and the purple lines are individual regression functions, bounded by each participant’s maximum cursor error in both directions. (Inset) Group mean regression weights (β). Error bars = 1 s.e.m.

We developed several models to account for the observed biases. The *gating* model meant to capture a process whereby sensorimotor error signals act to limit, or “gate,” the updating of value representations during reinforcement learning. Thus, in the gating model, movement errors act as a cue that directly modulates how value updating proceeds. To implement this, we allowed value updates to occur at different rates following hit and miss trials,

$$V_{t+1}(a) = \begin{cases} V_t(a) + \eta_{hit}\delta_t, & \text{if } hit_t \\ V_t(a) + \eta_{miss}\delta_t, & \text{if } miss_t \end{cases} \quad (2.3)$$

where $V_t(a)$ is the value of target a at trial t (Equation 2.1), η is the learning rate, and δ is the reward prediction error (Equation 2.2). By having two learning rates, the model can differently scale the reward prediction error based on the trial outcome: When η_{miss} is low relative to η_{hit} , the estimate of the target’s value changes minimally after a miss trial,

in effect offloading “credit” for the failed outcome from the target to a different source (e.g., movement error). When η_{miss} is high relative to η_{hit} , the value is updated at a relatively faster rate, assigning credit for the failed outcome to the target.

In the *probability* model, the probability of reward is explicitly represented and incorporated into learning. In this model, the motor error is still acting as a contextual cue that affects value updating, but in a manner that suggests participants are explicitly tracking the chance that they hit the target. To implement this, the probability of reward and the magnitude of reward (“payoff”) were separately tracked,

$$\delta_{prob,t} = r_t^* - \hat{p}_t(a) \quad (2.4)$$

$$\hat{p}_{t+1}(a) = \hat{p}_t(a) + \eta_{prob}\delta_{prob,t} \quad (2.5)$$

$$E_{t+1}(a) = \begin{cases} E_t(a) + \eta_{payoff}\delta_{payoff,t}, & \text{if } hit_t \\ E_t(a), & \text{if } miss_t \end{cases} \quad (2.6)$$

$$V_{t+1}(a) = \hat{p}_{t+1}(a)E_{t+1}(a) \quad (2.7)$$

where $\hat{p}_t(a)$ is the agent’s estimate of the probability that target x will yield a hit on trial t , and $E_t(a)$ is the expected payoff magnitude from target a on trial t . Separate learning rates (η_{prob} , and η_{payoff}) update distinct reward prediction errors, where $\delta_{payoff,t}$ is equivalent to the standard reward prediction error used in the gating model, and $\delta_{prob,t}$ is a prediction error that uses a binary outcome r_t^* that takes on a value of 1 or 0 based on whether trial t was a hit or miss, respectively. Critically, payoff estimates (E) are only updated after hits. Payoff (E) and probability ($\hat{p}_t$) terms are multiplied to yield the total value V .

We also included a “basic TD” model, which had a single learning rate for all outcomes (Equation 2.1), as well as a “subjective utility” model, which modeled if and how participants valued different point values in a nonlinear fashion (Niv et al., 2012). The subjective utility model is identical to the basic TD model with one additional free parameter, α , which

exponentially transforms observed rewards r into subjective rewards R , which are used for value updating:

$$R_t = r_t^\alpha \quad (2.8)$$

Lastly, all target values (V s) were transformed into choice probabilities using the softmax action selection function:

$$P(a) = \frac{e^{V(a)\beta}}{e^{V(a)\beta} + e^{V(b)\beta}} \quad (2.9)$$

where β constitutes the inverse temperature parameter. Higher values of β lead to less decision noise (i.e. “stickier” preferences), and *vice versa* for lower values of β .

All models were fit using the observed trial outcomes and choices. Free parameters were optimized to minimize the negative log posterior between the modeled and real choice data using the MATLAB function *fmincon*. Free parameters (η) were optimized individually for each participant, and β was optimized as a single value across all participants. The latter constraint was imposed to limit the interaction between β and η , as these parameters are correlated. The values V for each target were initialized to 0 for all fits. Learning rates (η) were bounded between 0 and 1, and inverse temperature (β) was bounded between 0.05 and 10. Model fits were evaluated using the Bayesian information criterion and models were compared by both summing the BIC values within each condition and performing t-tests between BIC values across models.

Both the gating and probability models fit the data best, with a slight advantage for the probability model overall (Figure 2.3A). Summed BICs show a better fit for the gating model in the Spatial condition, and better fits for the probability model in both the Binary and Standard conditions. We also performed t-tests between each model’s BIC values for each subject in each condition: There were no significant differences between the gating and

probability model BICs in the Spatial ($p = 0.88$) or Binary conditions ($p = 0.10$), but the probability model provided a significantly better fit in the Standard condition ($p < 0.01$).

The gating model parameter values suggest that credit assignment could occur via a lower learning rate after motor errors relative to hits (Figure 2.3B, solid line) – in other words, motor errors in the spatial condition may act as a “cue” which limits value updating. We note here that what we have called “low” learning rate values for the gating model are actually well within normal learning rate values for reinforcement learning studies (Niv et al., 2012; Gershman, 2015). Thus, it is the difference between η_{miss} and η_{hit} within conditions that could be used to interpret the observed biases. One interpretation here is that while miss trials still appear to “punish” the selected target in the Spatial condition (i.e. $\eta_{miss} > 0$), the resulting update is smaller, relative to the hit trials, than in the other conditions.

The probability model parameters suggest another interpretation: First, all conditions appear to be tracking both probability and payoff (Figure 2.3C, solid line), but those values affect value updates at different rates relative to each other. In the Standard condition, $\eta_{probability}$ is high relative to η_{payoff} , which means changes in estimated hit probabilities (unexpected hits or misses) strongly alter the modeled probability values, creating volatile probability tracking. In the Spatial condition, $\eta_{probability}$ is low relative to η_{payoff} , suggesting a stable estimate of probability, but an unstable estimate of reward. This latter result could suggest that, in the Spatial condition, when a reward is lower than expected, the chosen target is immediately strongly punished; however, a target is not punished as harshly or quickly on miss trials, consistent with the results of the gating model.

We note that the current experiment does not strongly support one model over the other; critically, however, both models imply that the effect of reward prediction errors on learning is modulated in the context of salient motor errors: In the gating model, this comes about via a lower η_{miss} parameter versus the η_{hit} parameter in the Spatial condition (Figure 2.3B); in the probability model, this modulation is built into the model, because the classic reward

prediction error, $\delta_{payoff,t}$, is only computed on trials *without* a salient motor error (via the conditional statement in Equation 2.6). Thus, in terms of a neural account of the results, both models predict a diminished effect of negative reward prediction errors on choice in the event of motor errors (see section 2.3).

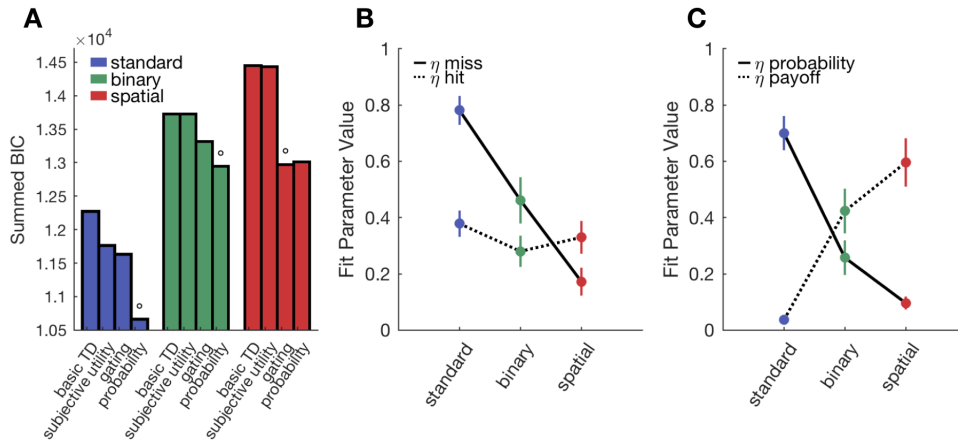


Figure 2.3: Modeling results. (A) Summed BIC values for each model and experimental condition. Lower values imply better fits. The gating and probability models outperformed both the subjective utility and basic TD model in all conditions. Open circles specify the best-fitting model in each condition. Best-fitting models were, according to the BIC metric, significantly more likely to explain the data than the next-best-fitting model. (B) Parameter values for the gating model, indicating learning rates for value updating following miss (black line) or hit (grey line) trials. (C) Parameter values for the probability model, indicating learning rates for value probability updating (black line) and payoff magnitude (grey line). Error bars = 1 s.e.m.

In an additional experiment, we investigated the following issue: How is behavior in this task modulated when the action execution system is compromised? Individuals with cerebellar degeneration often exhibit impairments in sensorimotor adaptation, a deficit that has been attributed to an inability to encode and/or use sensory prediction errors (Tseng et al., 2007; Shadmehr et al., 2010; Schlerf et al., 2012). We hypothesized that this impairment affects participants' choice biases, with the specific prediction that it would reverse the motor-error-related effects seen in the healthy controls (i.e. the bias toward the HMHR target in the Spatial condition).

To examine this hypothesis, individuals with cerebellar degeneration and age-matched

controls were recruited for a third experiment, where they were tested in a counterbalanced within-subject design on the three conditions shown in Figure 2.1. Reach adaptation was observed in both groups, and adaptation was marginally attenuated in the cerebellar group ($p = 0.09$). In terms of choice biases, the control group replicated, within subject, the results of the initial experiment (Figure 2.4). In contrast, the patient group showed a consistent bias for the low miss probability/low reward target (LMLR) across all three conditions (Figure 2.4; main effect of condition: $p < 0.01$; main effect of group: $p < 0.05$). Based on our *a priori* hypothesis, a planned direct comparison of the Spatial condition, which had the most salient motor errors, revealed a significant group difference ($p < 0.05$). This could suggest that patients treated negative reward prediction errors on Spatial miss trials similar to standard negative reward prediction errors. Whether this is the result of differences in perceived motor competence between the groups, or a more direct neural mechanism, is discussed in section 2.3.

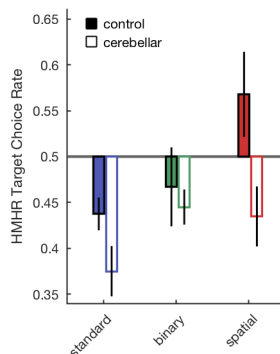


Figure 2.4: Neuropsychology results. Choice behavior for participants with cerebellar degeneration and age-matched controls in the three response conditions: Standard (blue), Binary (green), and Spatial (red). Error bars = 1 s.e.m.

The results outlined in this section suggest that a) motor errors act as a cue that affects how simple decisions are made, b) in some cases these errors appear to be discounted, suggesting that the RL system does not update choices based on the erroneous implementation of those choices, and c) if the motor system is compromised (via cerebellar damage), this discounting effect is diminished. Next, we adjusted the task to address this question within

a single response context (reaching), rather than between multiple response contexts. Our hypothesis was that the observed results represent a trial-by-trial, phasic process that can be revealed within-subject. The goal was to eventually optimize the task for fMRI scanning.

2.2 Multiplexed error signals within a single task context

The results of section 2.1 show how the RL system might selectively discount motor error trials. This was tested between response modalities (i.e. button presses vs. reaches). This leaves open an important question: Can we differentiate between motor errors and standard negative reward prediction errors within a single context?

The following experiment was conducted to address one important limitation of the previous study: The design of the Spatial condition in the previous study conflates reward probability with motor error – a non-rewarded trial is always presented as a motor error. This conflation makes it impossible to separate how participants react specifically to whether or not they hit the target, versus whether they received a reward or not.

To address this conflation, we designed a task that has three possible outcomes on each trial: hitting the target and being rewarded (Rew+), hitting the target and not being rewarded (Rew-), and missing the target (Miss). By removing this conflation we can get a better handle on unique effects of motor errors on reinforcement learning signals in the brain. In this section I first present behavioral results of this updated task, and in the section following it I present the results of an fMRI investigation using the task.

2.2.1 Task design and method

Participants held a digitizing pen and made reaching movements across a digitizing tablet (Wacom; sampling rate 100 Hz). Stimuli were displayed on a 17-inch LCD monitor, horizontally mounted 25 cm above the tablet. The task was controlled by custom software written in MATLAB (MathWorks), using the Psychophysics Toolbox (Brainard, 1997). The targets

(four circles, 1 cm diameter) were displayed in a rectangular arrangement about the start point, each 10 cm from the start position (Figure 2.5). At the start of each trial, the participant was guided to the start position by a visual cursor when the hand was within 10 cm of the start location. After maintaining that position for 500 ms, a “Go” cue signaled the participant to reach to the selected target. We imposed similar time constraints, boundary constraints, and cursor perturbations as the experiments in section 2.1.

However, several important things differentiate this task from earlier versions, namely that 1) trial outcomes were determined by fixed probabilities on each trial, and 2) trials had three possible outcomes (Figure 2.5, right): Rew+ (cursor hits target, participant is rewarded), “Rew-” (cursor hits target, no reward is received), and “Miss” (cursor misses target, no reward is received). Each target had a probability of 0.4 for a Rew+ trials, and mixed probabilities for Rew- vs. Miss outcomes, respectively, 0.1/0.5, 0.2/0.4, 0.4/0.2, and 0.5/0.1. The assignment of these probabilities was counterbalanced across target locations.

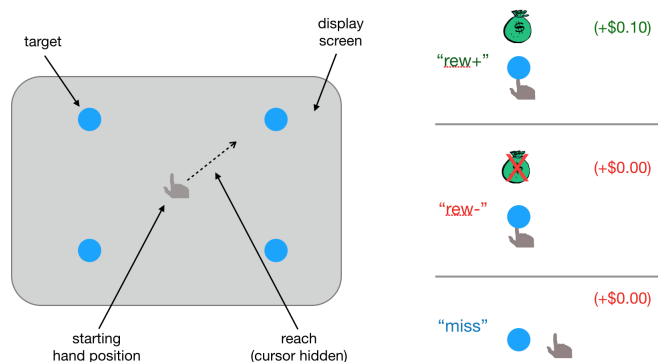


Figure 2.5: Task design. Participants selected targets using a graphical tablet, with vision of the hand occluded. Unlike the previous study, three outcomes were possible on each trial: hitting the target and receiving reward (Rew+), hitting the target but not receiving reward (Rew-), and missing the target. Cursor feedback was given on all trials, similar to the Spatial condition of the previous study.

2.2.2 Behavioral results

We analyzed participants' behavior based on whether they stayed or switched their target choice on a given trial, dependent on the outcome they received on the previous trial. Trial-by-trial switching provides a more detailed way to look at bias than the between-subject “high-miss-high-reward target bias” used in section 2.1.

Similar to the previous study, intriguing behavioral biases were observed, even though no particular strategies earned more average reward than any others: Participants were less likely to switch to a new target after miss trials than any other outcome (Figure 2.6; all p 's < 0.01).

This bias replicates our previous results (Figure 2.2A, spatial group), suggesting that motor errors may act as a cue that the decision outcome had an intrinsic cause, thus limiting updating of that choice's value. With this method established, the next step was to test our neural predictions: Our primary hypothesis was that reward prediction error signals in the associative striatum (NAc) would not appear to code for outcomes that were the result of motor errors – that is, action execution errors would not lead to updating of the chosen target's value, and thus no reliable negative reward prediction error would be detected.

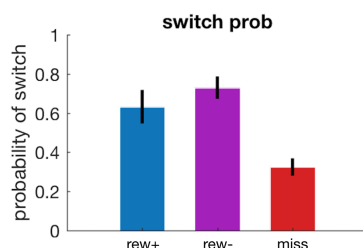


Figure 2.6: Behavioral results. Participants were most likely to switch following a Rew- trial, followed closely by Rew+ trials. On Miss trials, however, participants were more likely to repeat their selection, consistent with previous results. Error bars = 1 s.e.m.

2.3 fMRI investigation of striatal prediction errors in the context of action execution

The mechanism of reinforcement learning (RL) has been well characterized at a computational and neural level: Cortico-basal-ganglia loops (Alexander et al., 1986; Haber, 2003) leverage dopamine-dependent plasticity to guide value learning by slowly integrating trial-by-trial reward history (Equations 2.1 and 2.2; Sutton and Barto, 1998; Schultz, 1997). Dopaminergic neurons in the midbrain signal the requisite reward prediction errors (Montague et al., 1996). These neurons have projections to the striatum, allowing it to update action value estimates using prediction errors (Barto, 1995; Reynolds et al., 2001; Samejima et al., 2005). This results in improved value estimates, which can eventually lead to optimized selection of actions by cortical circuits (Daw and Doya, 2006; Bogacz and Gurney, 2007).

Recent evidence from studies in human and animal models have updated the model described above, suggesting that prediction errors in the striatum reflect both “model-free” and “model-based” teaching signals (Daw et al., 2011), where the former represents a classic running average of current states and scalar rewards (Equation 2.1), and the latter involves computing prediction errors based on a mental simulation of future states (Langdon et al., 2017).

Complicating the story further, prediction errors in the human striatum appear to be modulated by various cognitive factors, including attention (Leong, Radulescu, et al., 2017), episodic memory (Wimmer et al., 2014; Bornstein et al., 2017), and working memory (Collins et al., 2017). These results imply that prediction errors in the striatum are “smarter” than previously thought, integrating various sources of contextual information. In relation to the present study, since actions and choices are often conflated in studies of reinforcement learning, it is not known if the successful motor execution of decisions influences the representation of prediction errors in the striatum. The behavioral and modeling results presented

in sections 2.1 and 2.2 suggest that negative reward prediction errors should be disrupted, or “bypassed” in the event of execution errors, a neural hypothesis which is tested here.

2.3.1 Task design and method

The task was similar to that presented in section 2.2: On each trial, three stimuli were presented, depicted as cartoon slot machines, and the participant was required to choose one (Figure 2.7). The participant was instructed that each stimulus had some probability of yielding a reward and that they should try and earn as much money as possible. No explicit information was given to indicate that these probabilities differed between the stimuli. Critically, the participant was told that each trial was independent of the next (i.e. that the outcome on trial $t-1$ did not influence the outcome on trial t), and that they had a fixed number of trials in the task over which to maximize their earnings.

The participants in the current study were required to indicate their decisions by making a wrist movement with the right hand to the desired stimulus. The movement was performed by moving a wooden dowel (held with a pen grip) across an MRI-compatible drawing tablet. To initiate each trial, the participant moved her hand into a start area, which corresponded to the center of the tablet and visual display. The start area was displayed as a hollow white circle (radius 0.75 cm) and a message, “Go to Start”, was displayed until the hand had reached the start position. To assist the participant in finding the start position, a white feedback cursor (radius 0.25 cm) corresponding to the hand position was visible when the pen was within 4 cm of the start circle. As soon as the cursor entered the start circle, the three choice stimuli were displayed along with the text “Wait”. The three choice stimuli were cartoons of slot machines (0.7 cm by 0.7 cm). Each was presented at the identical three locations for all trials, displayed along an invisible ring (radius 4.0 cm) at 30°, 150°, and 270° relative to the origin.

After an exponentially determined jitter ($\mu = 1$ s, truncated range = 1.5 s - 6 s), the

“Wait” text was replaced with the message “GO!” (Figure 2.7). Reaction time was computed as the interval between the appearance of the go signal and the moment when the participant’s hand left the area corresponding to the start circle. The participant had 2 s to begin the reach; if the RT was greater than 2 s, the trial was aborted and the message “Too Slow” appeared. Once initiated, a reach was considered complete when the radial amplitude of the movement reached 4 cm, the distance to the invisible ring. This point also defined the movement time (MT) interval. If the MT exceeded 1 s, the trial was aborted and the message “Reach Faster” was displayed.

The feedback cursor was turned off during the entirety of the reach, and reappeared as a small, hand-shaped cursor (dimensions: 0.35 cm X 0.35 cm) at the end of the reach, displayed along the invisible ring. The actual position of this feedback cursor was occasionally controlled by the experimenter (see below), although the participant was led to believe that it corresponded to their hand position at 4 cm (similar to the experiments in 2.1 and 2.2). To help maintain this belief, the trial was aborted if the reach was $> \pm 25^\circ$ away from any one of the three stimuli, and the message “Please Reach Closer” was displayed. On trials in which the reach terminated within the required temporal and spatial boundaries, the cursor feedback remained on the screen for 1.5 s, and the participant was instructed to maintain the final hand position during this period. At the end of the feedback period, the entire display screen was blanked and replaced by a fixation cross presented at the center for an exponentially jittered inter-trial interval ($\mu = 1$ s, truncated range = 2 - 8 s).

Trial outcomes (Figure 2.7) matched those in section 2.2 (Figure 2.5): On Rew+ trials, the feedback cursor was accompanied by the appearance of a small money-bag cartoon above the chosen stimulus and \$0.10 would be added to the participant’s monetary total. On Rew- trials, the feedback cursor was accompanied by the same moneybag but with an overlaid red “X,” and no money was added to the participant’s total. Participants were informed at the start of the experiment that, as with real-life slot machines, sometimes there is a payoff

and sometimes there is no payoff, and the Rew+ and Rew- outcomes corresponded to these two possibilities. The third outcome consisted of “Miss” trials, in which the feedback cursor would appear outside the chosen stimulus. No money bag was presented on these trials and the monetary total remained unchanged, as in Rew- trials. Participants were informed at the start of the experiment that, like Rew- trials, no reward would be earned on trials in which their reach failed to hit the chosen target.

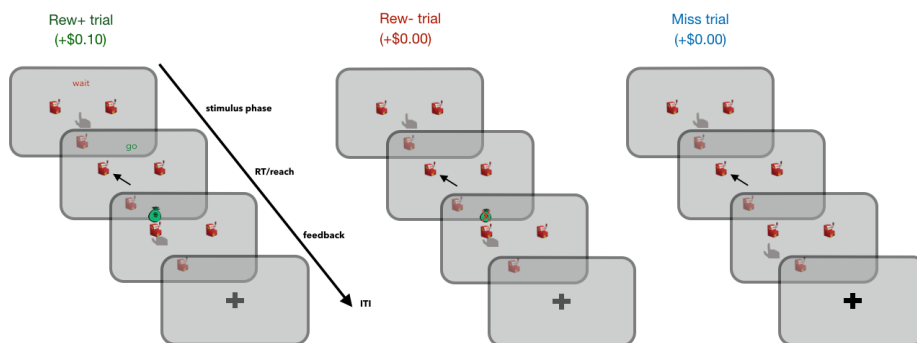


Figure 2.7: Task design. Participants selected one of three slot machines on each trial by reaching to them using a digital tablet in the fMRI scanner. Three trial outcomes were possible: On Rew+ trials (A), the cursor was seen as hitting the target and a reward was received; on Rew- trials (B), the cursor also hit the target but no reward was received; on Miss trials (C), the cursor was seen landing outside the target and no reward was received.

For the latter two outcomes, the feedback clearly distinguished between trials in which the movement was signaled as accurate but the slot machine failed to provide a payout (Rew-), and trials in which the movement was signaled as inaccurate and no reward was given due to the execution error (Miss).

Like the task in section 2.2, for all three targets the probability of obtaining a reward (Rew+) was 0.4. They differed in their ratio of Rew- and Miss probabilities, with the three targets assigned to one of the following ratios for Rew-/Miss outcomes: 0.5/0.1, 0.3/0.3, and 0.1/0.5. Thus, the targets varied in terms of how likely they were to result in execution errors. The positions of the stimuli assigned to the three Rew-/Miss probability ratios were counterbalanced across participants.

By fixing the outcome probabilities, we introduce two important features in our task design: First, given that the Rew+ outcome was the same for all three targets, there is no optimal choice behavior; that is, the participants should earn the same total bonus (on average) regardless of their choices. This constraint also means that the range of trial-by-trial reward prediction errors (RPEs; see below) should not strongly change over the course of the experimental session, as positive/negative outcome probabilities are fixed.

Second, similar to the tasks in 2.1 and 2.2, to maintain fixed probabilities for each target we had to vary whether the cursor feedback was veridical on a trial-by-trial basis. If the movement missed the target ($> \pm 3^\circ$ from the center of the target) on Rew+ and Rew- trials, the cursor was depicted to land within the target. If the movement hit the target on a Miss trial, then the cursor was depicted to land outside the target, and the size of the displacement on Miss trials was drawn from a skewed normal distribution ($\mu = 19 \pm 2.3^\circ$), which was truncated to not be less than 3° (the target hit threshold) or greater than 25° (the criterion required for a valid reach). Given the difficulty of the reaching task in the scanner and the strict temporal and spatial constraints, we expected that participants would be unaware of the feedback manipulation (see Results). The visual display was projected on a mirror attached to the MRI head coil. The participants were familiarized with the task during the anatomical scan and performed 30 practice trials for which they were not financially rewarded. Participants received a post-experiment questionnaire at the end of the task to assay their awareness of the feedback perturbations.

Trials were excluded from the analysis if the reach was initiated too slowly (i.e. RT was too long; $0.4 \pm 0.7\%$ of trials), completed too slowly (i.e. MT was too long; $2.4 \pm 4.5\%$ of trials), or terminated out of bounds (i.e. reach was too inaccurate; $1.2 \pm 2.0\%$ of trials). For the remaining data, we first evaluated participants' choice biases: For each target, the choice bias was computed by dividing the number of times the participant chose that target by the total number of choice trials. Second, we looked at switching biases: These were

computed as the probability that the participant switched to a different target on trial t given the outcome of trial $t-1$ (Rew+, Rew-, or Miss). An additional switching analysis was conducted based on only the reward outcome of trial $t-1$ (i.e. rewarded versus non-rewarded trials) by collapsing Rew- and Miss trials together. One-sample t -tests were used to evaluate if differences in choice and switching biases deviated significantly from zero.

To further evaluate potential predictors of switching, a logistic regression (Scholl et al., 2015) was conducted using choice switching on trial t as the outcome variable (1 for switch, 0 for stay). Seven predictors were entered into the regression: 1) The reward outcome experienced by the participant on trial $t-1$ (1 for reward, 0 for no reward), 2) the movement execution outcome experienced by the participant on trial $t-1$ (1 for a hit, 0 for a miss), 3) the experimentally fixed Rew- to Miss trial probability ratio of the chosen target on trial t , 4) the absolute cursor error magnitude experienced by the participant on trial $t-1$, 5) the veridicality of the feedback on trial $t-1$ (1 for veridical feedback, 0 for perturbed feedback), 6) the interaction of absolute error magnitude X the veridicality of the feedback on trial $t-1$, and 7) the current trial number. The multiple logistic regression was computed using the MATLAB function `glmfit`, with a logit link function. All regressors were normalized (for plotting purposes). One-sample t -tests were used to test for significant regression weights across the sample.

We also analyzed how movement feedback altered reaching behavior. In particular, we were interested in whether participants were sensitive to the non-veridical feedback provided on trials requiring a perturbed cursor position, and incorporated such feedback into changes in their movements. This analysis was identical to the one depicted in Figure 2.2C.

2.3.2 fMRI methods

Whole-brain imaging was conducted on a 3T Siemens PRISMA scanner, using a 64-channel head coil. MRI-optimized pillows were placed about the participant's head to minimize

head motion. To start the scanning session, anatomical images were collected using a high-resolution T1-weighted MPRAGE pulse sequence (1 x 1 x 1 mm voxel size). During task performance, functional images were collected using a gradient echo T2*-weighted EPI sequence with BOLD contrast (TR = 2000 ms, TE = 28 ms, flip angle = 90°, 3 x 3 x 3 mm voxel size; 36 interleaved axial slices). A field map was acquired to improve registration and limit image distortion from field inhomogeneities. For one participant, a field map was not collected.

Functional data were collected in a single run that lasted approximately 40 min. For one participant, the run was split into two parts due to a brief failure of the drawing tablet. Because of the self-paced nature of the reaching task (i.e., variable time taken to return to the start position for each trial, reach, etc.), the actual time of the run, and thus number of total TRs, varied across participants. The run was terminated once the participant had completed all 300 trials of the task.

Preprocessing and data analysis were performed using FSL v. 5.0 (FMRIB) and SPM12. Given the movement demands of the task and length of the scanning run, multiple steps were taken to assess and minimize movement artifacts. After manual skull-stripping using FSL's brain extraction tool (BET), we registered the functional images to MNI coordinate space using a rigid-body affine transformation (FLIRT) and applied the field map correction, and spatially smoothed the functional data with a Gaussian kernel (8 mm FWHM). We then applied the independent components motion-correction algorithm ICA-AROMA (Pruim et al., 2015) to the functional data to identify and remove components identified as head-motion artifacts. As a final preprocessing step, we applied standard motion correction (MCFLIRT) and temporal filtering with a 100 s high-pass filter. Four participants were excluded from further analyses due to excessive (> 3 mm pitch, roll, or yaw) head motion.

Four GLMs were performed. For the first three GLMs, we imposed a family-wise error cluster-corrected threshold of $p < 0.05$ (FSL FLAME 1), with a cluster-forming threshold of

$p < 0.001$. Task-based regressors were convolved with the canonical hemodynamic response function (double Gamma), and the six motion parameters were included as regressors of no interest.

The first GLM was designed to functionally define ROIs that were sensitive to reward. Trial outcome regressors for the three trial types (Rew+, Rew-, Miss) were modeled as delta functions concurrent with visual presentation of the trial outcome. Task regressors of no interest included boxcar functions that spanned both the wait period and reach period. The contrast, $\text{Rew+} > (\text{Rew- and Miss})$ was performed to identify reward-sensitive ROIs. Resulting ROIs were visualized, extracted, and binarized using the xjview package for SPM (<http://www.alivelearn.net/xjview>). Beta weights were extracted from the resulting ROIs using FSL's *featquery* function. To identify areas sensitive to visuomotor motor errors while controlling for reward, this GLM was used for a second contrast: Miss trials $>$ Rew- trials.

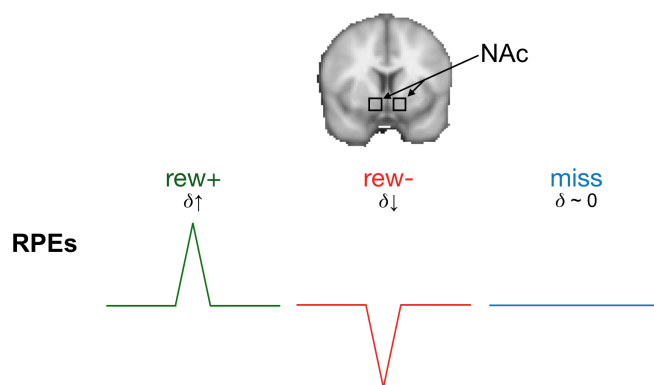


Figure 2.8: Predictions (*NAc* = nucleus accumbens; *RPE* = reward prediction error)

A second whole brain GLM was used to measure striatal RPEs (Figure 2.8). Three separate parametric RPE regressors, corresponding to RPEs for each outcome, were entered into the GLM to account for variance in trial-by-trial activity not captured by the three binary outcome regressors. Beta weights for each RPE regressor were extracted from the striatum ROI determined by the first GLM using FSL's *featquery* function. Nuisance regressors included the wait period, reach period, and the three outcome regressors.

The third GLM was designed to identify brain areas parametrically sensitive to motor execution errors. The regressors here were limited to Miss trials and included a single separate parametric absolute cursor error regressor, which tracked the magnitude of angular cursor errors on Miss trials. Nuisance regressors included the wait period, reach period, and the three outcome regressors.

The fourth GLM was an exploratory psychophysical interaction (PPI) analysis (Friston et al., 1997). In a PPI, a task-specific regressor and ROI time course regressor are included into the same model with the critical addition of a third regressor that models the interaction between the other two regressors, capturing variance in activity not singularly attributable to either regressor alone. A mean time series from the striatum ROI was extracted using `fslmaths` and added (unconvolved) to the model as an additional regressor. Interaction regressors between the striatum time course and the three individual outcome regressors were also included. Nuisance regressors included the wait period, reach period, and the three outcome regressors. We imposed a family-wise error cluster-corrected threshold of $p < 0.05$ (FSL FLAME 1), with a relaxed cluster-forming threshold of $p < 0.05$ (see Results). All voxel locations are reported in MNI coordinates, and all results are displayed on the averaged MNI brain.

2.3.3 Behavioral results

Choice data revealed no consistent biases across the three stimuli (Figure 2.9A, all p 's > 0.15). We note that, unlike in our previous study (section 2.1; McDougle et al., 2016), the probability and magnitude of reward on each trial were stable over the course of the task for each stimulus.

Critically, trial-by-trial switching behavior offers a more detailed way to look at bias (Figure 2.9B). Consistent with previous results, participants were more likely to switch to a different stimulus following Rew- trials compared to Miss trials ($p < 0.001$). Moreover, they

were more likely to switch after Rew- trials compared to Rew+ trials ($p < 0.001$), and showed no difference in switching rate after Rew+ and Miss trials ($p = 0.45$). Overall, participants were, on average, more likely to switch following a non-rewarded trial (Rew- or Miss) than a rewarded one ($p < 0.001$; Figure 2.9B inset), suggesting that they were generally sensitive to receiving a monetary reward. We note, however, that switch rates were still relatively high after Rew+ trials, suggesting that participants may have been at least somewhat sensitive to the fact that every target had an identical reward probability.

In sum, the switching behavior indicates that participants responded more negatively to Rew- outcomes compared to Miss outcomes, even though both yielded identical economic results. This finding is consistent with the hypothesis that the effect of negative RPEs on choices following execution errors is attenuated relative to successful execution, consistent with previous modeling efforts (Equations 2.3-2.9; McDougle et al., 2016).

A regression analysis was used to further probe switching behavior (Figure 2.9C). The first two regressors, reward and execution outcome, recapitulated the results shown in Figure 2B, where the reward outcome (reward vs. no reward) and the execution outcome (hitting the target vs. missing) both had a strong effect on switching behavior: Getting rewarded on trial $t-1$ negatively predicted switching on trial t (i.e., predicted staying over switching), reflecting the positive Rew+ trials (t-test for regression weight difference from 0: $p = 0.029$). In contrast, hitting the target on trial $t-1$ had a positive impact on the probability of switching on trial t , driven by the aversive Rew- trials ($p = 0.027$). Both effects were tempered by the Miss trials, which, as shown before, led to reduced switching (Figure 2B). Consistent with Figure 2.9A, the Rew-/Miss probability ratio of the selected target on trial t had only a marginal effect in the regression analysis ($p = 0.061$).

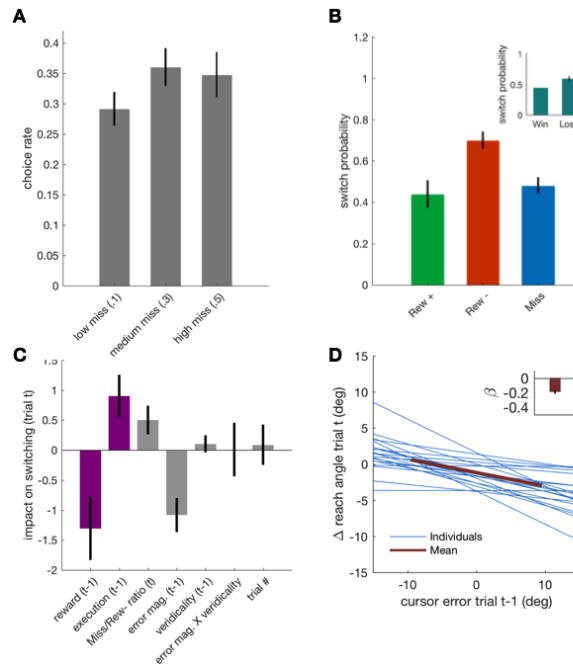


Figure 2.9: Behavioral results. (A) Participants’ biases to select stimuli with a different ratios of Rew- to Miss trials. (B) Average switch probabilities separated by the outcome on the previous trial. Inset: switch probabilities separated by rewarded trials (Rew+) versus unrewarded trials (Rew- and Miss collapsed). (C) Logistic regression on switch behavior. (D) Logistic regression on change in reach angle as a function of signed cursor errors on the previous trial. This analysis is limited to trials in which participant’s reach on trial t-1 was accurate, but the cursor was perturbed away from the target (Miss trial). Inset: average regression weight. Error bars = 1 s.e.m.

Interestingly, the absolute magnitude of the cursor error on trial t-1 negatively predicted switching on trial t; that is, after relatively large errors, participants were more likely to repeat the same choice again ($p = 0.002$). This effect did not appear to be driven by the veridicality of the error, as neither the regressor for the veridicality of feedback, nor the interaction between veridicality and error magnitude, predicted switching ($p = 0.49$ and $p = 0.98$, respectively). Lastly, switching behavior did not predictably fluctuate over the duration of the experiment (trial number regressor; $p = 0.80$).

Overall, we had to perturb the cursor position on 58.4% of trials. Most of these (47.6%) were “false hits,” where the feedback cursor was moved into the target region following an actual miss. The other 10.8% of the perturbed trials were false misses, in which the cursor was displayed outside the target following an actual hit.

We had designed the Miss-trial perturbations with the goal of keeping the participants unaware of the feedback perturbations, while also providing large, visually salient execution errors. Critically, the mean size of the perturbed Miss trial errors was 11.20° larger than veridical Miss trial errors ($p < 0.001$), raising the possibility that participants could be made aware of the perturbations. The results from a post-experiment questionnaire were equivocal: When asked if the feedback was altered, the mean response on a 7-point scale was 4.3, where 1 is “Very confident cursor accuracy is fully controlled by me,” and 7 is “Very confident cursor accuracy is partially controlled by me.” However, it is not clear if the question itself biased participant’s answers, so further analyses were needed.

As noted above, in terms of switching, the regression analysis indicated that participants responded similarly to trials following veridical or perturbed cursor feedback (Figure 2.9C). We next examined if adjustments in reaching direction were responsive to non-veridical errors as they would be expected to after veridical errors. To this end, we analyzed trial pairs in which the same stimulus was chosen on two consecutive trials where the first reach had been accurate but resulted in a false miss (mean number of pairs per participant = 18.4). If participants “believe” the perturbed feedback, the second movement should be shifted in the opposite direction of the preceding perturbation. We note that while we could perform the same analysis following veridical misses or perturbed hits, a shift would be expected simply from regression to the mean (mean hand angle from target = 0.12°), whereas in this case, the hand would be shifting away from the mean.

Consistent with this prediction, a regression analysis showed that heading direction did indeed shift in the opposite direction of the perturbation on the subsequent trial ($t(19) = -6.36, p < 0.001$; Figure 2.9D). This could be interpreted as resulting from sensorimotor adaptation (Krakauer, 2009), however, the size of the trial-by-trial adjustment was larger than that typically observed in studies characterizing implicit sensorimotor adaptation (Wei and Kording, 2010). As such, we assume it primarily reflects explicit adjustment in aiming,

perhaps happening in concert with implicit adaptation (Mazzoni and Krakauer, 2006; Taylor et al., 2014; see next chapter). Taken together, both the regression and movement analyses, and to a lesser extent the questionnaire, indicate that manipulation of the cursor feedback did not have a significant impact on participants' choice behavior.

2.3.4 Modeling results

The modeling approach was similar to that described in Equations 2.1-2.9, with several important modifications. First, all models included decay parameter γ (Collins et al., 2014), which drives the three values, V , toward their initial values on each trial (initialized at 1/the number of actions, or 1/3):

$$V = V + \gamma(1/3 - V) \quad (2.10)$$

This parameter is especially important here as every trial has identical outcome probabilities assigned to each target, and thus there is no optimal choice that should be maintained over trials.

Our previous results showed that participants discount Miss trials, suggesting a tendency to stick with a given choice following execution errors (sections 2.1-2.2; McDougale et al., 2016). However, it is not known if this tendency reflects behavior based on prediction error computations or arises from a different mechanism. To model an additional route to Miss discounting, we included a “persistence” parameter, ϕ , in the *softmax* function that produces the probability of each choice (P), with inverse temperature β ,

$$P = \frac{e^{\text{miss_prev}(\phi * \text{choice_prev}) + V\beta}}{\sum e^{\text{miss_prev}(\phi * \text{choice_prev}) + V\beta}} \quad (2.11)$$

where `miss_prev` and `choice_prev` are indicator vectors, indicating, respectively, whether the previous trial was a Miss (1 for Miss, 0 for Rew+/Rew-) and which action was chosen. If

ϕ is positive, the learner is more likely to repeat a choice after a Miss trial; if ϕ is negative, she is more likely to switch. This parameter was meant to represent a bias factor that was distinct from that resulting from the computation of the value update (Bornstein et al., 2017), and we thought this would improve our model fits.

We started with two control, or “Standard” models: In the Standard(2η) model, distinct learning rates, η , were included to account for updating following negative reward prediction errors (unrewarded trials) and positive reward prediction errors (rewarded trials),

$$V_{t+1}(a) = \begin{cases} V_t(a) + \eta_{Rew+}\delta_t, & \text{if } Rew+_t \\ V_t(a) + \eta_{Rew-,Miss}\delta_t, & \text{if } Rew-_t, Miss_t \end{cases} \quad (2.12)$$

where η_{Rew+} and $\eta_{Rew-,Miss}$ are the learning rates for updates following Rew+ or Miss/Rew- trials, respectively. Allowing positive and negative RPEs to update values at different rates has been shown to provide better fits to human behavior compared to models in which a single learning rate is applied after all trials (Niv et al., 2012; Gershman, 2015). We also included a second variant of this model, the Standard(no- ϕ) model, that was identical to the Standard(2η) model but did not include the ϕ parameter, in case that parameter was unnecessary. Both Standard models acted as control models, where motor errors were not treated as special outcomes.

Next, we attempted to recreate the two winning models (Gating and Probability) from our previous study (section 2.1). The Gating model is identical to the model described in Equation 2.3, with one key difference: It has unique learning rates for each of the *three* possible trial outcomes (Rew+, Rew-, and Miss). Thus, the Gating model allows for values to be updated at a different rate following execution errors (Miss) or selection errors (Rew-).

Lastly, the Probability model, like the one presented in section 2.1 (Equations 2.4-2.7), separately tracks the probability of successful execution for each target (E) and the likelihood ($\hat{p}$) of receiving a reward if execution was successful. Both the Gating and Probability models

included the decay (γ) and persistence (ϕ) parameters.

All models were fit using the observed trial outcomes and choices (Figure 2.10). Free parameters were optimized to maximize the log posterior probability of the observed choice data given the model, using the MATLAB function *fmincon*. Models were fit separately for each participant. All learning rate parameters were constrained to be between -1 and 1. Negative values were allowed given that we did not have an *a priori* reason to assume η_{Miss} would be positive, and thus opted to be consistent across all learning rate parameters and models. The persistence parameter (ϕ) was constrained to be between -5 and 5, and the decay parameter (γ) was constrained to be between 0 and 1. The inverse temperature parameter (β) was constrained to be between 0 and 100, and a Gamma(2,3) prior distribution was used to discourage extreme values (Leong, Radulescu et al., 2017). The fitting procedure was conducted 100 times for each model using different randomized starting parameter values to avoid local minima during optimization, and the resulting best fit was used in further analyses. Model fits were evaluated using both the Bayesian and Akaike information criteria.

All models predicted choice behavior better than chance – average per-trial likelihoods of each model predicting participants' true choices were significantly greater than the chance value of 0.33 (all p 's < 0.001; Figure 2.10C). The best model was the Gating model: First, the Gating model provided the best fit compared to the other three models in terms of both BIC and AIC (all p 's < 0.001, Figure 2.10A, B). Second, the Gating model had a higher average per-trial likelihood of predicting choices over the next best model ($t(19) = 4.61, p < 0.001$; Figure 2.10C). Third, the Gating model provided the best fit for 16 of the 20 of the participants (Figure 2.10D). Consistent with our previous results (section 2.1; McDougale et al., 2016a), the modeling analysis indicates that in tasks that allow for motor execution errors, a dedicated update parameter for such trials strongly improves the model fit, suggesting that motor errors act as a special “class” of outcome.

We next examined the estimated parameter values for the Gating model. The learning

rates on Miss trials, η_{Miss} , and Rew- trials, η_{Rew-} , were both greater than zero ($t(19) = 2.89, p = 0.009$ and $t(19) = 2.64, p = 0.016$, respectively), with η_{Miss} showing the lowest median value (medians: $\eta_{Miss} = 0.07$, $\eta_{Rew+} = 0.13$, $\eta_{Rew-} = 0.23$). Parameter values were not normally distributed, and a Wilcoxon sign-rank test revealed no significant difference between η_{Miss} and η_{Rew-} ($p = 0.62$). Lastly, the persistence parameter (ϕ) in the Gating model came out significantly greater than zero ($t(19) = 2.49, p = 0.022$). This observation suggests that the observed high behavioral persistence after Miss trials may be facilitated by an additional mechanism outside of value updating (Bornstein et al., 2017).

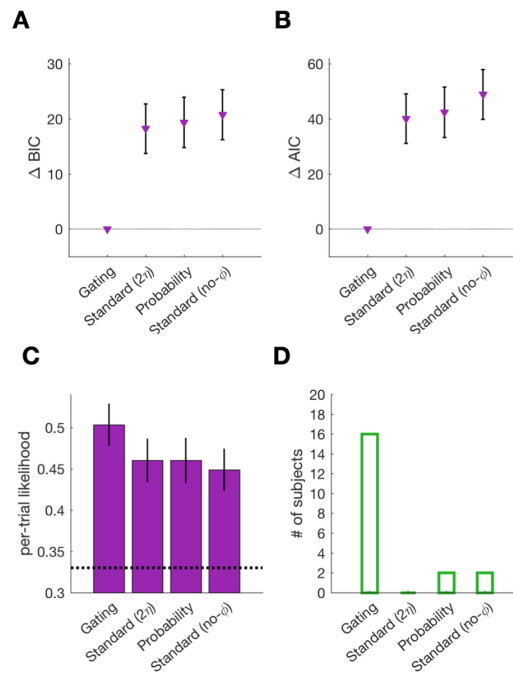


Figure 2.10: Modeling results. (A) Bayesian information criterion (BIC) and (B) Akaike information criterion (AIC) comparisons of each model, referenced to the BIC and AIC values of the Gating model (higher values represent weaker fits). (C) Average per-trial likelihoods of each model predicting participants' true choices, with a chance value of 0.33. (D) Number of participants best-fit by each model (using AIC). Error bars = 1 s.e.m.

The Gating model has six free parameters, raising a concern about over-parameterization. To address this, we simulated (300 times) choice data using the random parameter values (within the constraints) and then fit the Gating model to the simulated data to confirm

that we could recover those values. We could recover the six free parameters with varying degrees of precision. For the three learning rates (η_{Rew+} , η_{Rew-} , η_{Miss}) we achieved r-values of, respectively, 0.82, 0.79, and 0.77 for the correlation between simulated and recovered values. The persistence parameter (ϕ) was recovered with an r-value of 0.76. However, the inverse temperature (β) and decay (γ) parameters were less recoverable, at r-values of 0.54 and 0.56.

Previous studies have shown that movements toward high value choices are more vigorous (i.e., faster) compared to low value choices (Reppert et al., 2015; Seo et al., 2012; Niv et al., 2007). Given that we used reaching movements in the current study, we could ask if this phenomenon is observed in the current context, looking at the effect of modeled values on both reaction time (RT) and movement time (MT). Overall, reaction times were moderately fast ($\mu = 0.59 \pm 0.13$ s) and movement times were quite fast ($\mu = 0.13 \pm 0.06$ s). MT and RT, as well as the modeled values (V) of selected choices (from the Gating model), were extracted for each participant, de-trended using linear regression (due to gradual decreases in RT and value time courses), and z-scored. Linear regressions were performed to quantify the influence of relative values on trial-by-trial MTs and RTs. Consistent with previous results on movement vigor and value, model-derived values negatively predicted MT (regression beta value relative to 0: $t(19) = -3.28, p = 0.004$); that is, higher-value choices were accompanied by faster movements (i.e. shorter movement times; Figure 2.11). No significant relationship was observed between RT and model-derived values ($t(19) = 0.38, p = 0.71$). We speculate that this null result may be a function of the design of the task (Figure 2.7), which included an enforced wait period before movement.

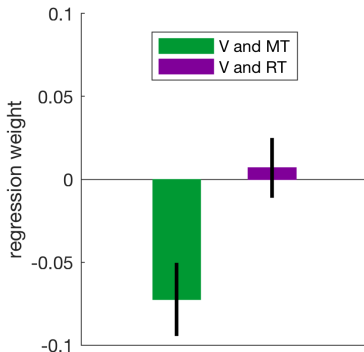


Figure 2.11: Modeled choice value and movement vigor (V = modeled value; MT = movement time; RT = reaction time). Error bars = 1 s.e.m.

2.3.5 Imaging results

Figure 2.12 and Table 1 show the results of the whole-brain contrasts for motor error processing (Miss > Rew-), restricted to trials in which reward was withheld, and reward processing (Rew+ > Rew- and Miss). For the motor error contrast, three broad clusters were revealed (green), including a single elongated cluster spanning bilateral premotor cortex (PMC), supplementary motor area (SMA), and the anterior division of the cingulate (ACC), as well as two distinct clusters in both the left and right inferior parietal lobule (IPL). This pattern is consistent with previous work on cortical responses to salient motor errors (Krakauer et al., 2004; Diedrichsen et al. 2005; Bedard and Sanes, 2014; Seidler et al., 2014).

The reward contrast (red/yellow) revealed four significant clusters: one spanning bilateral striatum, one spanning bilateral ventromedial prefrontal cortex (vmPFC), one spanning bilateral posterior cingulate (PCC), and a single cluster in left orbital frontal cortex (OFC). These ROIs are broadly consistent with areas commonly associated with reward (McClure et al., 2004; Schultz, 2015).

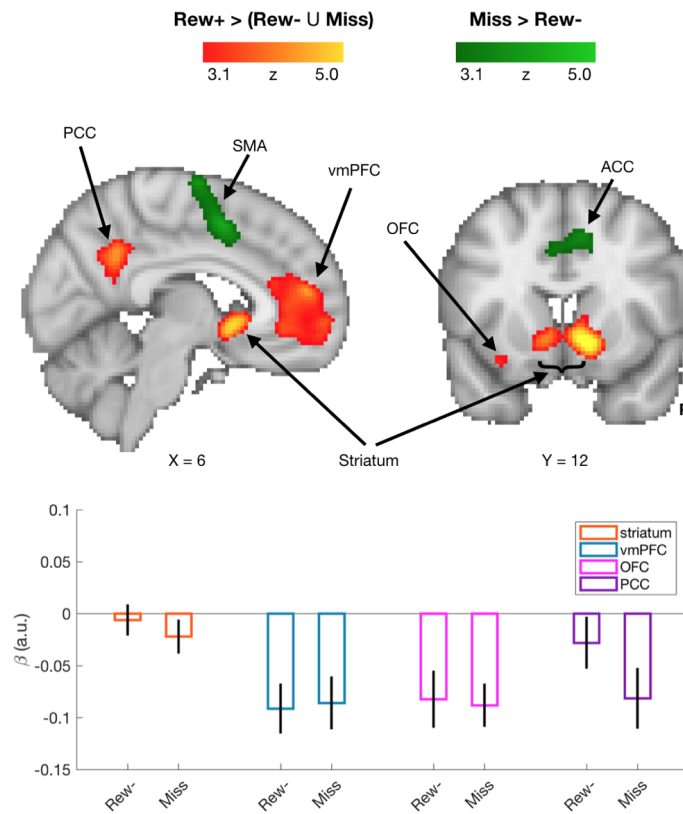


Figure 2.12: Trial Outcome Contrasts. Top: Results of whole-brain contrasts for Rew+ trials > Rew- or Miss trials (red/yellow), and Miss > Rew- trials (green). For the motor error contrast (green), three significant clusters were revealed, with a single cluster spanning bilateral premotor cortex, supplementary motor area (SMA), and the anterior division of the cingulate (ACC), as well as two distinct clusters in both the left and right inferior parietal lobule. In the reward contrast (red/yellow), four significant clusters were revealed, in bilateral striatum, ventromedial prefrontal cortex (vmPFC), left orbital-frontal cortex (OFC), and posterior cingulate cortex (PCC). Bottom: Beta weights extracted from each reward contrast ROI for the (orthogonal) Rew- and Miss trial outcomes. Error bars = 1 s.e.m.

Examination of feedback-locked betas on Rew- and Miss trials should identify gross differences in activity in these ROIs distinct from the more fine-grained parametric RPE modulations to be explored in the model-based analysis (see below). Directly comparing the two negative outcome trial types revealed that average activity in the four ROIs was similar for Rew- and Miss trials, with no significant differences seen in the striatum ($t(19) = 0.88, p = 0.39$), vmPFC ($t(19) = -0.24, p = 0.81$), nor OFC ($t(19) = 0.25, p = 0.81$), and a marginal difference in the PCC ($t(19) = 1.95, p = 0.07$).

Analysis/Region	x (mm)	y (mm)	z (mm)	# voxels
Rew+ > (Rew- U Miss)				
striatum	5	11	-6	691
vmPFC	0	46	2	2710
L OFC	-37	33	-12	567
PCC	-1	-51	32	911
Miss > Rew-				
SMA/PMC/ACC	10	-2	61	1733
R IPL	60	-24	33	1395
L IPL	-55	-25	26	733
PPI (Miss X Striatum)				
L IFG/L putamen	-40	20	2	944
Error Size (Miss)				
M1/PMC	0	-23	55	10399
V1/R Cb	15	-80	9	2704
R LOC/R IPL	-55	-25	26	568

Table 1: Significant Clusters. All clusters survived cluster correction at the $p < 0.05$ level (FLAME 1) with cluster-forming threshold of $p < 0.001$, with the exception of the PPI analysis, which used a threshold of $p < 0.05$. Coordinates are in MNI space and correspond to the cluster's center of gravity. vmPFC = ventromedial prefrontal cortex; OFC = orbitofrontal cortex; PCC = posterior cingulate cortex; SMA = supplementary motor area; ACC = anterior cingulate cortex; IPL = inferior parietal lobule; IFG = inferior frontal gyrus; M1 = primary motor cortex; PMC = premotor cortex; V1 = primary visual cortex; Cb = cerebellum; LOC = lateral occipital cortex

Although we failed to observe differences in the contrast of Rew- and Miss trials in the initial imaging analysis, the large behavioral differences (Figure 2.9B) between these two outcomes suggests that a more granular analysis was required. In our second GLM, separate parametric RPE regressors for the three possible trial outcomes were constructed by convolving trial-by-trial RPE values derived from the Gating model with the canonical hemodynamic response function (HRF). Beta weights for the three regressors were then extracted from the striatum ROI delineated by the first GLM. The key results are depicted in Figure 2.13: As predicted, striatal activity parametrically tracked trial-by-trial RPEs following Rew+ trials ($t(19) = 3.26, p = 0.004$) and Rew- trials ($t(19) = 2.76, p = 0.013$).

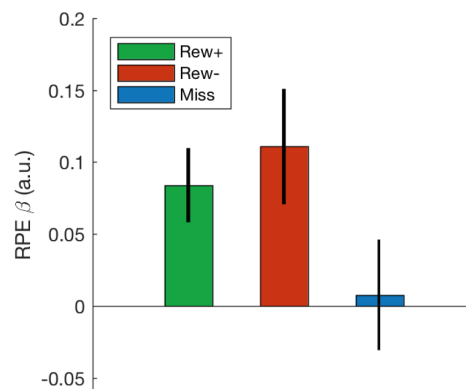


Figure 2.13: Outcome coding in the striatum: Average reward prediction error (RPE) beta weights within the striatum ROI for each trial outcome type. Error bars = 1 s.e.m.

Strikingly, striatal activity did not appear to encode RPEs following Miss trials ($t(19) = 0.20, p = 0.84$). Moreover, the strength of RPE coding was significantly greater on Rew- trials than on Miss trials ($t(19) = 2.52, p = 0.020$), marginally greater on Rew+ trials than on Miss trials ($t(19) = 1.84, p = 0.081$), and not significantly different between Rew+ and Rew- trials ($t(19) = -0.74, p = 0.47$). Consistent with the gating hypothesis (section 2.1; McDougle et al., 2016), these results suggest that striatal coding of RPEs is attenuated following motor execution errors. A consequence of this would be that choice value updating would be effectively bypassed, contributing to the observed behavioral biases (sections 2.1, 2.2, Figure 2.9).

A GLM analysis was also conducted to confirm that observed execution errors were processed in predicted motor-related areas. The absolute error size on Miss trials was entered as a parametric regressor in a whole brain analysis. Consistent with previous research (Seidler et al., 2014), error magnitude was correlated with the modulation of the BOLD response in anterior cingulate cortex, dorsal premotor cortex, dorsal cerebellum (lobule VI), and primary visual cortex (Table 1). No significant voxels in the striatum were identified in this analysis.

To investigate areas that may act in concert with the striatum to evaluate action execution cues, we performed an exploratory psychophysiological interaction (PPI) connectivity

analysis. Our PPI analysis quantifies correlations in BOLD activity between the striatal ROI and other brain areas that are more pronounced during Miss trials relative to other trial outcomes (Friston et al., 1997). For this analysis, we relaxed our cluster-forming threshold to $p < 0.05$. The PPI revealed a significant functional interaction on Miss trials between the striatal ROI and an elongated cluster that consisted of, primarily, left inferior frontal gyrus (IFG) and left putamen (Figure 2.14). As a point of comparison, we performed a similar PPI analysis on Rew+ and Rew- trials. Here, no significant clusters were found between the striatal ROI time course and the rest of the brain. We note that, although the FSL FLAME algorithm used in all group-level analyses limits false positive rate relative to most other approaches (Eklund et al., 2016), the results displayed in Figure 2.14 were not significant at more conservative statistical thresholds, and thus should be viewed with appropriate caution.

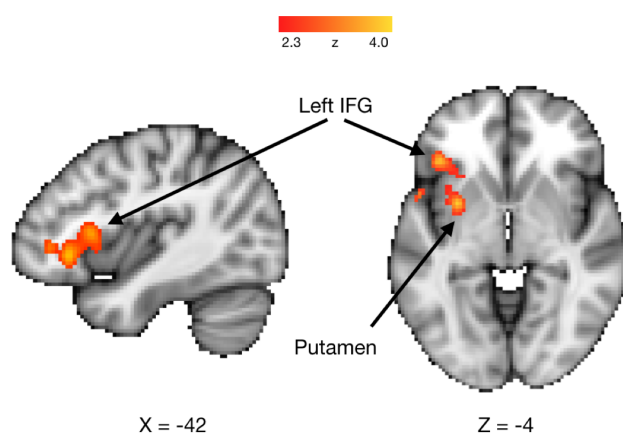


Figure 2.14: PPI Analysis. Activity in left inferior frontal gyrus (IFG) and left putamen was correlated with activity in the striatum ROI on Miss trials

2.4 General discussion

In natural settings, outcomes frequently depend on the organism's ability to execute a coordinated action: A hungry osprey not only faces the challenge of detecting the silhouette of a striped bass but must accurately execute her plan of attack. Previous studies have examined

the impact of motor noise on choice behavior (Wu et al., 2009), focusing on how representations of motor competence influence subjective utility functions. However, such descriptive accounts do not provide a mechanism concerning how motor errors influence representations of extrinsic value. The studies presented in this chapter suggest one such mechanism – when execution of a selected choice is not successful, negative reward prediction errors in the striatum are attenuated. Therefore, in the settings studies here, motor errors do not affect future valuation of the chosen option.

Central to this proposed mechanism is the notion that this specific event – an error in motor execution – is communicated to the reinforcement learning system. These findings fit into a broader reevaluation of the nature of reward prediction errors in the mesostriatal dopamine system. Mounting evidence suggests that the striatum does not just compute a model-free prediction error, but is affected by high-level cognitive states, concerning, for instance, model-based predictions of future rewards (Daw et al., 2011; Langdon et al., 2017), sampling from episodic memory (Wimmer et al., 2014; Bornstein et al., 2017), top-down attention to relevant task dimensions (Leong, Radulescu et al., 2017), and the holding of stimulus-response relationships in working memory (Collins et al., 2017). We believe the present results add to this body of evidence, showing that cues concerning the implementation of a decision affect if and how the represented value of that decision is updated by a prediction error.

We note that this putative gating phenomenon in the striatum was not categorical; indeed, participants displayed varying degrees of gating both behaviorally and neurally (Figure 2.9B, Figure 2.13). One speculation could be that gating is a function of how optimistic a participant is that they could correct a motor error in the future — here, gating is useful only if one is confident in their execution ability, and are thus likely to persist with a decision until they hone their execution and successfully glean reward information about the selected stimulus. On the other hand, if one is not confident in their intrinsic execution abilities,

they may have a strong reward prediction error upon an execution error because they are pessimistic in their ability to implement a given decision in the future.

This interpretation could explain the curious result in the cerebellar patient group (Figure 2.4): Participants with degeneration of the cerebellum, which results in problems with both motor learning and motor execution, showed diminished “gating” behavior; that is, they avoided decisions that were difficult to execute, even at the cost of larger rewards. One more speculative hypothesis we put forward was that the cerebellum may be an important structure in a putative gating mechanism — perhaps communicating sensory prediction errors (Schlerf et al., 2012) to the basal ganglia via established bidirectional connections (Chen et al., 2014) — and, further, that this interaction could be observed through an fMRI connectivity metric. However, significant cerebellar activity only survived statistical correction in our analysis of cursor error size (Table 1), and the results of our planned analyses on trial outcomes did not reveal significant interactions between the cerebellum and striatum. Thus, we do not provide evidence that cerebellar error signals are affecting choice behavior in this kind of task. Another interpretation is that participants implement knowledge about the causal structure of the task to guide their decisions. Recent work has shown that if participants are told they are in control of a probabilistic outcome (but, similar to our task, are in fact not) they tend to alter how they estimate the probabilities of those outcomes (Green et al., 2010). Speculatively, if one’s motor “competency” is compromised (as it is in the cerebellar patient group), one could consider motor error a state of the world that is under limited volitional control. This could explain the different biases we observed in that group (Figure 2.4).

Via reverse inference, the results of our PPI connectivity analysis (Figure 2.14) are broadly consistent with putative behavioral roles for the left inferior frontal gyrus (IFG), suggesting that the left IFG is one candidate region involved in the observed gating of RPEs. Recent work supports the hypothesis that the left IFG inhibits belief updating fol-

lowing certain negative outcomes (Sharot et al., 2011; Sharot et al., 2012; Moutsiana et al., 2015), which is appealingly similar to the gating phenomenon presented here. Others have highlighted a more general role for the left IFG in controlled retrieval processes that apply goal-relevant knowledge in a top-down fashion (Badre and Wagner, 2007). We speculate that a perceived execution error could be interpreted as a specific case of a more generalized cue about the current situation the participant is in, where the specific consequence of this negative outcome is to inhibit value updating.

Overall, the results of this chapter offer one example of how action execution influences action selection. Specifically, errors in execution can provide an important contextual cue to upstream decision-making systems, informing them that the implementation of their action was flawed, not necessarily the selection of that course of action. The ecological value of this mechanism is clear – it is important for an organism to properly assign credit for outcomes to either selection or execution. In the next chapter, I take a different approach to studying selection vs. execution: How does the action selection influence action execution?

3 Selection informs execution: controlled selection and implicit execution processes in motor learning

Many classic studies of sensorimotor learning have been based on reaching and throwing movements, and the results have helped us to gain fundamental insights into foundational ideas such as the trade-off of speed and accuracy (Fitts, 1954) and internal representations of sensorimotor dynamics (Shadmehr and Mussa-Ivaldi, 1994). One important subfield of motor learning research employs adaptation tasks to ask how an internal model, a representation of body–environment interactions, is calibrated to support feedback and feedforward control (Wolpert et al., 1995). The internal model concept has provided a useful theoretical tool to understand how people adjust their motor execution when a visuomotor mapping is altered or perturbed.

Such sensorimotor adaptation tasks have been used to characterize processes responsible for calibrating the mapping between desired outcomes (selected actions) and motor commands (Krakauer et al., 2000). Previous research has focused on how this form of error-based learning proceeds in an implicit and automatic manner. However, recent work has revealed the operation of multiple learning processes, even in this simple form of learning (Taylor et al., 2014; Bond and Taylor, 2015; McDougle et al., 2015). This section of the thesis highlights the robust contribution of volitional, cognitive motor action selection to sensorimotor learning, and models how selection sets the stage for the adaptation of an implicit internal model.

In section 3.1 we develop and test a model that extends a previous “2-state” model of sensorimotor learning (Smith et al., 2006), suggesting that a putative “fast” learning process represents volitional action selection, and a putative “slow” learning process represents implicit recalibration of an internal model for reaching. In section 3.2 we extend this model to explain how selection and execution may interact. In our model, the selected motor plan

tags which population of units in the internal model get updated after errors, revealing an ongoing dynamic where selection and execution inform each other during learning.

3.1 Measuring explicit and implicit motor learning

The physics of the body and environment are in a continuous state of flux: not only do long-term changes arise from growth, development, and injury, but, in the short-term, muscles fatigue and sensory conditions fluctuate. The motor system must rapidly adjust to these variable conditions, and the ease with which we maintain calibration belies its computational complexity (Shadmehr et al., 2010).

To study this calibration process, researchers have employed a variety of learning tasks, including prism adaptation (Weiner et al., 1983; Martin et al., 1996; von Helmholtz, 1924; Redding et al., 2005), visuomotor rotations or “VMR” (Cunningham, 1989; Krakauer et al., 2000), and force-field learning (Shadmehr and Mussa-Ivaldi, 1994), in which a perturbation is introduced to alter the relationship between a movement and the resulting sensory feedback. Across a range of contexts, performance typically follows a stereotypical learning function driven by a gradient-descent-like process in which the error is reduced in a continuous, monotonic manner. When the perturbation is removed a persistent “aftereffect” is observed, taken as the signature of a recalibrated sensorimotor mapping. Over time, the aftereffect diminishes at roughly the same rate as that observed during the initial acquisition phase, eventually returning to the baseline, non-adapted state. Figure 3.1 shows a simulated canonical motor learning curve in the context of adaptation to a rotational perturbation (i.e. the VMR task).

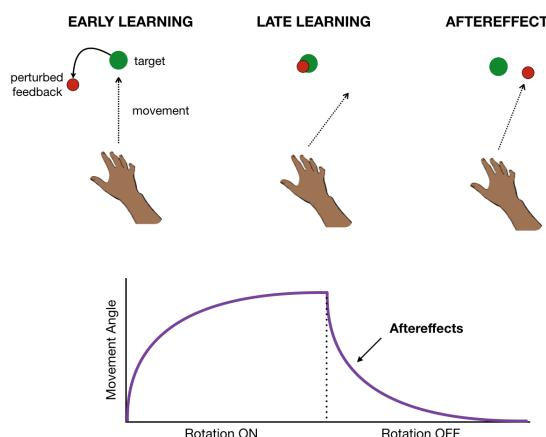


Figure 3.1: The motor adaptation curve

However, this narrow formulation misses a common-sense approach to the problem faced by the participants in such experiments: While throwing darts one evening, imagine that you don a pair of fashionable prism glasses, which make your dart land far to the right of the target. It would be reasonable to suppose that an intelligent agent would take steps to volitionally compensate for this perturbation. For example, you might aim to the left of the target on the next trial. Indeed, such compensatory strategies are essential on windy days for golfers and placekickers.

In one oft-cited prism adaptation example, some individuals displayed extremely rapid learning, completely reducing their error on a throwing task after a single trial (Martin et al., 1996). When queried, these individuals could explicitly report a strategy, whereby they estimated the error induced by the glasses and purposefully aiming in the opposite direction to negate the perturbation. Interestingly, this strategy proved to be unstable, with successive movements increasingly overcompensating for the prismatic distortion. When instructed to throw “where the target appeared,” their performance again took the form of the stereotypical learning function, indicating that performance may reflect the combined effects of strategy use and implicit recalibration.

One clever study that directly examined this hypothesis used a variant of the standard visuomotor rotation (VMR) task: participants were given explicit information about the

perturbation and were instructed to use a compensatory strategy (Mazzoni and Krakauer, 2006). Vision of the hand was occluded and feedback was limited to the display of a circle that indicated the position of the hand at the end of the movement. After an initial block of trials with veridical feedback, a 45° counterclockwise perturbation was imposed. Crucially, after two reaches in this altered environment the experimenter intervened, describing the perturbation and instructing the participants to aim in the clockwise direction. To facilitate the use of this strategy, landmarks were positioned at 45° intervals around the target. Thus, by aiming to the landmark 45° clockwise from the target, the perturbation could be fully negated.

As would be expected, participants performed perfectly on the subsequent trial: using an aiming strategy enabled one-trial learning. However, over the next 80 trials, the movements of the participants began to “drift” in the direction opposite the perturbation. This paradoxical behavior – where performance worsened with practice – suggests that the motor system continued to calibrate the motor commands based on the mismatch between the intended reach location (the aiming landmark) and visual feedback, while ignoring feedback about task accuracy (the difference between the target location and observed feedback). In a subsequent experiment, it was shown that the drift reversed with extended training, an effect attributed to an adjustment in the aiming strategy (Taylor and Ivry, 2011). This non-monotonicity, together with evidence using various other methods, has made it clear that strategic action selection and implicit recalibration constitute dissociable learning processes, which are summed in the observed task performance (Taylor et al., 2014).

To deconstruct this sum into its constituent parts, Taylor and colleagues (2014) developed a VMR task that provides a trial-by-trial measure of the contributions of explicit aiming and implicit recalibration. To assay strategic action selection, participants verbally reported their aim direction before each reach, providing these reports both before the perturbation and over the course of learning (Figure 3.2). Using a simple subtractive procedure (reach angle minus

aiming angle), one can estimate the precise state of implicit recalibration in a continuous manner. Interestingly, the lion's share of early learning, especially with large perturbations, is associated with aiming and not recalibration (Bond and Taylor, 2015). Moreover, and perhaps more surprising, aiming remains prevalent even in the late stages of learning, a result that challenges the standard belief that asymptotic performance only reflects a fully recalibrated sensorimotor mapping. Various control conditions indicate that these results are not an artifact of the aiming report task itself (Taylor et al., 2014): Critically, the learning curves and aftereffects in this task are similar to those observed in standard visuomotor rotation tasks, suggesting that strategic processes operate even when the task context (e.g., instructions, landmarks, etc.) does not prime their use. Lastly, the results above suggest that the two learning processes respond to different types of errors: Explicit selection is sensitive to performance error (Bond and Taylor, 2015), whereas implicit adaptation is sensitive to sensory prediction error, or the difference between expected and observed sensory outcomes (Mazzoni and Krakauer, 2006).

To further investigate these two learning processes, formalize their time courses, and characterize their interaction, we designed an experiment that allows us to model explicit selection strategies and implicit adaptation separately.

3.1.1 Task design and modeling methods

One common approach to teasing out multiple learning processes is to combine computational modeling and psychophysics. In a now classic study, Smith and colleagues (2006) applied this method to tasks where participants reached in a force-field. Normally in such tasks, learning curves are modeled as a single process. In a series of elegant experiments, the authors showed how learning of novel force field dynamics is actually driven by two concurrent learning components – one which is fast and flexible, and one which is slow and robust. The key experiment that revealed these processes looked at the phenomenon of spontaneous recovery

(also termed “rebound”).

In rebound paradigms, one force field is learned, it is briefly followed by an oppositely-signed force field, which is then followed by an extinction phase with no force fields and no visual errors (Smith et al., 2006). In this task, participants display a surprising behavior: Although they can rapidly learn the second force-field after being adapted to the first, they spontaneously “rebound” in the extinction block to the reaching dynamics they originally used to counter the first force-field. The only way to account for this odd discontinuity in the learning curve is to incorporate two learning components in your model, one which accounts for the rapid learning of the second force-field and one which is “stuck” in its first adapted state in the extinction phase (Smith et al., 2006).

This oft-cited model has been widely accepted in the motor learning literature. However, its underlying mechanisms are not well understood. Here we tested the hypothesis that the “fast” and “slow” learning components are homologous to, respectively, an explicitly-reportable action selection process and an implicit adaptation process. (Note: The findings in this section were recently published in a research article, McDougle et al., 2015).

Ten right-handed participants (mean age 21 years, range 18–33) were recruited from the research participation pool maintained by the Department of Psychology at Princeton University in exchange for course credit. The protocol was approved by the Princeton University’s Institutional Review Board and all participants provided informed consent.

In each study discussed in this Chapter, all participants made center-out, horizontal reaching movements to visually displayed targets (7 mm radius), moving their right hand across a digitizing tablet while holding onto a digitizing pen (Intuos Pro; Wacom). Movement trajectories were sampled at 100 Hz. The stimuli were displayed on a 17-inch LCD computer monitor (Dell) mounted horizontally 25 cm above the tablet. The monitor occluded vision of the hand and a small, circular cursor (3.5 mm radius) provided continuous online visual feedback during each reach. The task was controlled by custom software written in Python

(<http://www.python.org>).

At the start of each trial, participants were required to position their hand in a central starting position with the aid of a visual ring, which represented the distance between the hand and the starting position. After maintaining this position for 1 s, a green visual target appeared. After leaving the starting position, the participant's hand had to cross an invisible ring that contained the target in 500 ms to avoid a “too slow” warning delivered aurally to the participant automatically by the game software. If the center of the cursor landed within the target, a pleasant “chime” sounded; otherwise, a “buzz” sounded. Participants in all experiments were told to make sure to reach within the time limit on each trial and to try to land the cursor in the target on every attempt.

This updated sensorimotor adaptation task, depicted in Figure 3.2, ascertains the participant's explicit cognitive strategy on every trial by simply asking where they are aiming (Taylor et al., 2014). Thus, after learning, movements can be decomposed, simply through subtraction, into implicit and explicit components (Figure 3.2).

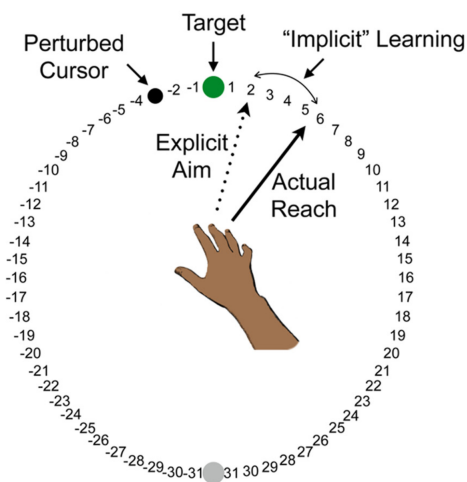


Figure 3.2: Task design. Participants verbally report, before each movement, where they planned to aim to make the cursor land on the target. The numbers are displayed around the full 360° circle. In all experiments, participant's vision of their hand was occluded.

We used the rebound paradigm described above, using a visuomotor rotation task (VMR) instead of a force-field task (Figure 3.3; Smith et al., 2006). The visual target appeared at

a single 0° location on every trial. The first 100 trials were baseline trials in which cursor feedback was veridical. On trials 101–300 (R1 epoch, blue), the cursor was rotated by 45° . On trials 301–320 (R2 epoch, green), the cursor was rotated by 45° in the opposite direction (i.e. counter-clockwise vs. clockwise). In trials 321–420 (EC epoch), a visual “error clamp” was placed on the cursor, making it move straight to the target regardless of the participant’s hand trajectory orthogonal to the target, thus abolishing visual errors. Numbered visual aiming landmarks were spaced 5.63° apart, and the numbers were arranged in a 360° circle (Figure 3.2). To assess task performance, we computed the heading angle of the reach as the angle midway along the trajectory (3.5 cm). Our modeling approach is outlined below.

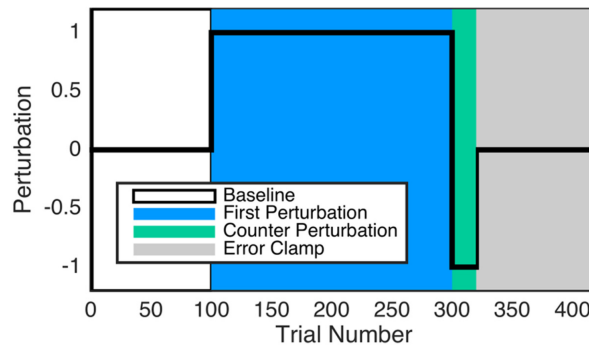


Figure 3.3: Perturbation schedule for all experiments, similar to Smith et al., (2006). Perturbations are normalized at 1/1 for illustrative purposes. There were no perturbations in the baseline block (white), perturbation one in the R1 counter-perturbation two in the R2 block (green), and error clamp in the EC block (gray).

We tested the hypothesis that the popular 2-state model of motor learning (Smith et al., 2006) could be explained by the concurrent operation of explicit learning (fast process) and implicit learning (slow process).

The standard state space model of adaptation (Thoroughman and Shadmehr, 2000) is defined as follows,

$$X_{t+1} = AX_t + B\varepsilon_t \quad (3.1)$$

$$\varepsilon_t = X_t - p_t \quad (3.2)$$

where ε is the error experienced at time t , defined as the difference between the current motor state X_t and perturbation p_t . A is the retention factor applied during the update of X , and B is the learning rate.

Smith et al. (2006) outlined a dual-process state space model that combines a fast component, denoted by an E in Equation 3.3, and a slow component, denoted by an I in equation 3.4, to yield the final output X (Equation 3.5). This model assigns separate retention factors and learning rates to each process, where $A^E < A^I$, and $B^E > B^I$.

$$X_{t+1}^E = A^E X_t^E + B^E \varepsilon_t \quad (3.3)$$

$$X_{t+1}^I = A^I X_t^I + B^I \varepsilon_t \quad (3.4)$$

$$X_{t+1} = X_{t+1}^E + X_{t+1}^I \quad (3.5)$$

Critically, the fast process learns quickly and is flexible, and the slow process learns slowly and is robust (Smith et al., 2006; Huberdeau et al., 2016). The model was fit to minimize the root-mean-square-error (rmse) between the model simulation and participant data using the *fmincon* function in MATLAB. The outputs of the fast (X^E) and slow (X^I) processes were fit to the time series of explicit and implicit learning, respectively. For fitting, the model was constrained such that $A^E < A^I$, and $B^E > B^I$. The stability of the model fits and the sensitivity of the initial conditions were tested using a grid search over 100 different starting parameter values. We found that the model fits were consistently highly stable within our constraints.

As mentioned above, instead of using a standard blocked design (i.e. baseline block, perturbation block, aftereffect block), we used a “rebound” paradigm (Figure 3.3). Rebound paradigms have been used to expose “latent” states of learning: In a rebound paradigm, participants first learn to counter a particular perturbation, then a second perturbation is briefly presented, and finally adaptation is measured in an extinction, or error clamp

phase. In most cases, participants learn the second perturbation easily but show “rebound” in the error clamp, reaching in a manner that suggests they are still adapted to the first perturbation. Such results provide evidence that multiple learning processes exist: a fast process that drives learning of the second perturbation, and a slow process that remains adapted to the first perturbation during the error clamp (Smith et al., 2006). Figure 3.4 shows a simulation of the two state model in our rebound paradigm design. As shown in that figure, rapid learning in the second perturbation phase (green) is the result of the fast process (solid black line), and rebound is a result of the slow process being “stuck” in the first state (dashed black line).

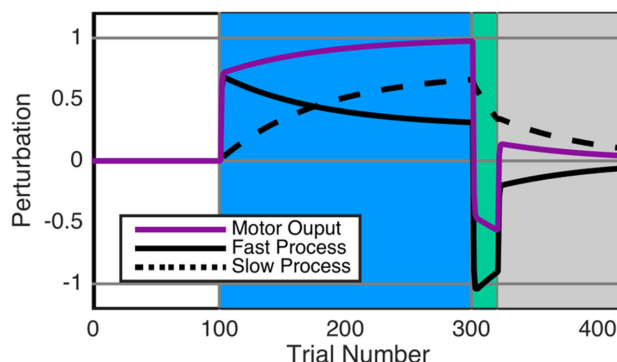


Figure 3.4: Model simulations. Simulations of the two-state model of sensorimotor learning. Note that the fast process (solid black line) learns at a fast rate but has low retention and the slow process (dashed black line) learns at a slow rate but has high retention. Motor output was defined as the combination of the fast and slow processes (purple line).

3.1.2 Behavioral and modeling results

Consistent with our hypothesis, we found a strong correspondence between the fast and slow states, and respectively, explicit and implicit learning processes (Figure 3.5). The fast process provided a strong fit to participants’ aiming data (i.e. their explicit motor plan), and the slow process provided a strong fit to the subtraction measure, which represents implicit adaptation of an internal model. Critically, the fast learning of the second rotation was clearly attributable to the explicit aiming process, and rebound was clearly attributable to

implicit adaptation (Figure 3.5).

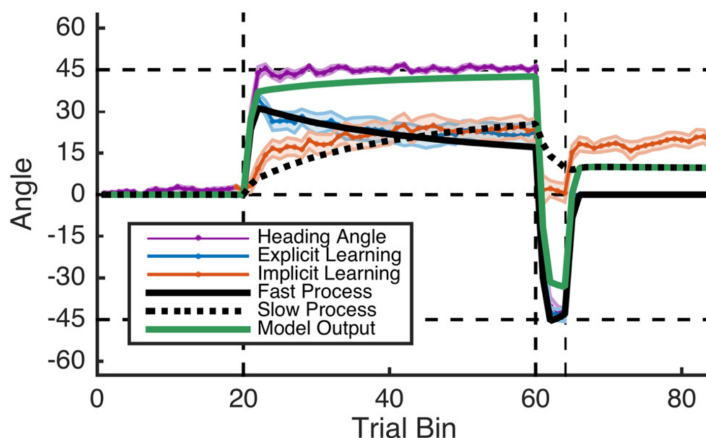


Figure 3.5: Behavioral results and model fit. The fast and slow processes of the model were fit individually to each participant’s respective explicit (blue) and implicit (red) learning data. Here, the model captures R1 behavior but fails to capture all implicit learning in the R2 epoch and undershoots the rebound in the EC epoch. Shading = 1 s.e.m.

We note, however, that the model fit of the slow process is far from perfect — indeed, in both the second perturbation block (trial bin 60) and the error clamp block (trial bin 64), the model fit is leaves much to be desired. In study 2.2, we attempt to resolve this issue.

3.2 Explicit action selection determines which movements are adapted

Having established the concurrent operation of explicit and implicit processes in visuomotor learning, our next question concerned the *interaction* of these processes. One starting point for our this investigation was the apparent “discontinuities” of the implicit learning time course in our single-target rebound experiment (Figure 3.5). Our hypothesis was that changes in action selection were also changing which implicit time course we were measuring; in effect, the implicit learning curve we observe may reflect only the state of adaptation of the currently-selected movement plan.

It is not well understood how explicit selection and implicit execution interact. Previous work shows that relative contributions of each process can be affected by perceptual factors;

for example, if the cursor feedback is continuous vs. a single endpoint sample, the relative ratio of implicit learning to explicit aiming will be higher (Taylor et al., 2014).

However, such findings do not suggest any particular communication between explicit selection and implicit adaptation – indeed, it has been shown that explicit learning merely picks up the “slack” when implicit learning is saturated: When comparing groups of participants who experienced different rotation sizes (e.g. 45° vs. 75°), implicit learning curves appear identical while explicit learning curves are much higher for the larger rotations (Bond and Taylor, 2015).

Given our hypothesis that action selection “tags” which movements get adapted, we reasoned that a generalization experiment would be the best test of this idea. By testing how participants moved to untrained locations, one can test if the generalization of adaptation is linked to selection. (Note: The findings in this section were recently published in a research article, McDougle et al., 2017).

The generalization of learning to novel situations plays a central role in perceptual, semantic, motor, and statistical learning, and has been aptly termed the first law of psychology (Shepard, 1987). In the motor domain, purely “local” learning would inevitably make the learner vulnerable to the curse of dimensionality: One will never make the same movement twice. Generalization in sensorimotor adaptation has been specifically defined: When a subject adapts their movements to counter a perturbation in a given movement direction, adaptation generalizes to adjacent directions following a Gaussian-like function (Thoroughman and Shadmehr, 2000; Poggio and Bizzi, 2004).

One important question concerns the “center” of the generalization function, which refers to the movement direction that displays the maximum degree of adaptation after learning (i.e. the first moment of the generalization function). Recent work (Day et al., 2016), using the explicit/implicit method described above, has suggested that the strongest candidate for the center of generalization is indeed the participant’s explicitly-selected plan: Adaptation-

related adjustments were not largest at the location of the learning target (target-based generalization), nor at the location that was most frequently traversed by the hand (movement-based generalization; Castro et al., 2011), but rather at the location the participant most frequently planned to move to achieve task success and be rewarded (Day et al., 2016).

To further investigate this, incorporate it into a model, and resolve the discontinuity of implicit learning in section 3.1, we conducted 4 experiments. First, we conducted a generalization experiment in an attempt to parameterize an average generalization function, while also replicating the results of Day et al (2016). In experiments 2-4 we test an updated state space model, which incorporates selection- (or “plan-”) based generalization, in three versions of the rebound paradigm.

3.2.1 Task design and methods

The method and apparatus was identical to that described in Chapter 3.2, with several modifications. The first 96 trials involved reaches to 16 target locations (0° , 22.5° , 45° , 67.5° , 90° , 112.5° , 135° , 157.5° , 180° , 157.5° , 135° , 112.5° , 90° , 67.5° , 45° , and 22.5°), which were presented in pseudorandomized blocks of 16 trials.

In the first 16 trials participants received online feedback to get accustomed to the reaching task, followed by 32 trials to get accustomed to endpoint feedback and then 48 trials in which participants received no feedback to measure reach biases at all 16 targets (Figure 3.6). Veridical endpoint feedback was restored for the next 16 trials, directed at a single target location (0° ; i.e., the target directly to the right of the start point), and participants were instructed to verbally report their aiming location on each trial before moving (baseline aim report block) so they could get accustomed to the reporting procedure. After this baseline block, for 96 trials (rotation block) participants experienced a 45° rotation (the rotation was -45° for the counterbalanced group) while reaching to a single learning target location (0°); participants were instructed to continue to report their aiming location on each trial. The

rotation block was designed to get participants to fully learn the perturbation at the single learning target and reach asymptote.

The following 180 trials constituted the generalization probe block, which was designed to test for generalization to novel target locations: On 50% of trials participants were exposed to 1 of the 15 probe target locations, while still reporting their aim but receiving no cursor feedback. Feedback was withheld at the generalization targets to restrict error-driven learning. On the other 50% of trials in the generalization block participants were exposed to the learning target and continued to receive rotated endpoint cursor feedback, to prevent decay of adaptation at the learning target. Learning trials and generalization trials were alternated to ensure a regular “top-up” of adaptation at the learning target. Finally, the last 48 trials constituted the aftereffect block, where participants were told to reach directly at each target, stopped reporting any aims, and received no feedback. Targets were pseudo-randomly presented in three blocks of 16 trials. The aftereffect block was designed to get a measure of the final level of implicit adaptation at each probe location.

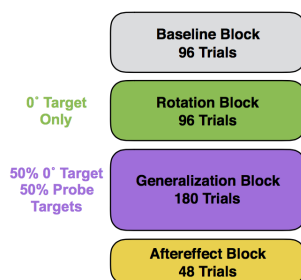


Figure 3.6: Generalization experiment. Text on left of colored blocks specifies what type of target (learning vs. probe) was presented in the respective block.

The mean implicit generalization function was fit with a Gaussian function. Each participant’s implicit generalization function was determined by calculating the mean implicit learning at each of the 16 locations during the probe trials. A Gaussian was fit to the group data via bootstrapping, resampling our subject pool with replacement 1000 times, and its parameters (height and width, with mean fixed at 0°) were optimized to provide the lowest

root mean square error (rmse) between the fitted curve and the bootstrapped mean function. Fits were optimized with the *fmincon* function in MATLAB.

3.2.2 Generalization results

Participants displayed the typical Gaussian generalization pattern: Implicit learning at the probe locations tapered off as a function of distance from the original learning location (Figure 3.7B, C). While it is not surprising that maximum adaptation is seen at the learning target, because only those trials had cursor feedback, the pattern shown in Figure 3.7 does not specify what aspect of behavior drives maximal implicit adaptation. The maximum adaptation observed at the learning target can reflect at least one of three variables: the learning target's location (target-based generalization), the movement made during learning trials (movement-based generalization), or the explicit plan during learning trials (plan-based generalization). Thus in Figure 3.7B the x-axis is conflated with both participants' movement and aiming directions during learning trials. This conflation is addressed later.

Figure 3.7B depicts the full generalization function of participants' movements (purple) and the individual generalization components for explicit (blue) and implicit (red) learning. As previously shown (Heuer and Hegele, 2011), explicit generalization was essentially flat across the workspace. Such a pattern could reflect true global generalization, a function of the landmarks and instructions in the task, or a combination of both; critically, our goal in designing the task was to induce global (flat) explicit generalization in order to unmask the “true” implicit generalization function (i.e., if aim directions were highly variable across probe trials, it would yield a highly distorted implicit generalization function). We note that the generalization of explicit learning in an unconstrained experimental setting may not be truly “flat” but would rather have its own nonuniform, perhaps also Gaussian, function. But for the purposes of the present study, we sought to isolate generalization of implicit learning by constraining explicit generalization.

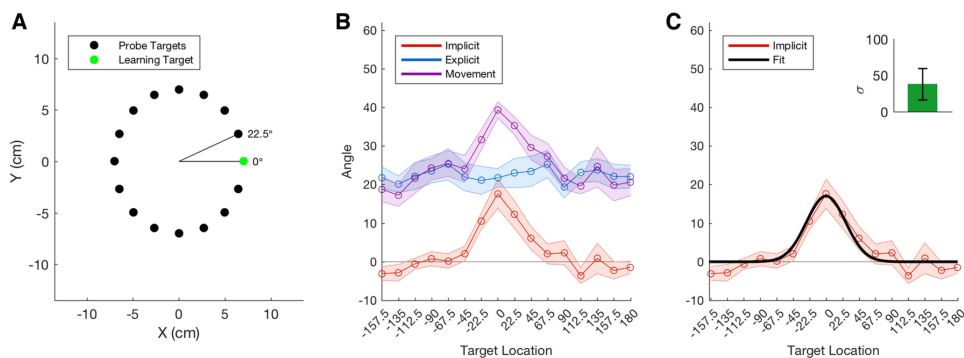


Figure 3.7: Generalization results. (A) Experimental setup. During the learning block (trials 97–208), participants were only presented with the learning target (0° target, green) and received rotated feedback. During the probe block, 50 percent of trials were rotated feedback trials at the 0° target, and 50 percent of trials were no-feedback trials at the probe locations (black). (B) Generalization functions are shown for participants’ hand angles (purple), aim reports (blue), and implicit learning (red). y-Axis represents the angular value of each learning component, at each target location, after being rotated to a common 0° axis for analysis. (C) A Gaussian was fit to the bootstrapped mean implicit learning function. Inset: the mean width parameter of the fit Gaussian function, with 95 percent confidence interval. Shading = 1 s.e.m.

Interestingly, global explicit learning causes an effective positive DC shift in the measured generalization function of movements (Figure 3.7B, purple). This DC shift is routinely observed in visuomotor generalization data (Krakauer et al., 2000; Taylor et al., 2013), and our results make the novel suggestion that such a shift may be a result of the additive generalization of explicit and implicit learning.

We determined the width of implicit generalization with a Gaussian function, via bootstrapping (see above); the average width was 37.76° (Figure 3.7C). This width value was used for all subsequent model simulations and fits. We note that the value is, qualitatively, in accordance with the width of previously described generalization functions (Krakauer et al., 2000; Mattar and Ostry, 2007; Taylor et al., 2013).

Our next analysis used the aftereffect block to determine the center of implicit generalization. In the aftereffect block, participants were told to aim directly to presented targets and all cursor feedback was removed, giving us a measure of implicit learning at each exact target location without the confound of an explicit plan toward a location away from the

target. If generalization is “target based,” the maximal aftereffects should be seen at the 0° target, which was the only target where participants received feedback during the task. If generalization is “movement-based” (Castro et al., 2011), the peak aftereffect should be seen around the 45° target, as the mean movement direction during learning block trials (0° target trials) was $42.80 \pm 0.50^\circ$. Finally, if generalization is plan-based, the peak aftereffect should be closest to the 22.5° target, as the mean aiming direction toward the learning target was $26.19 \pm 3.50^\circ$.

The generalization function during the aftereffect block is shown in Figure 3.8A. This function was created by taking participants’ mean reach directions to each target in the no-feedback aftereffect block and subtracting their mean reach direction to each target in the no-feedback baseline bias block. As predicted by plan-based generalization (Day et al., 2016), peak aftereffects were present near the 22.5° target. This result is especially compelling considering that no feedback was ever received during the rotation block at the 22.5° target trials, whereas at the 0° target location rotated feedback was consistently present through both the learning and probe blocks.

Importantly, since each subject had a unique mean aim direction during learning trials, we could perform a more quantitative analysis: First, a Gaussian was fit individually to each participant’s distribution of aftereffects, with free parameters for the height, offset, standard deviation, and mean value. The fitted mean value (i.e. the “center of generalization,” the Gaussian’s first moment) was labeled as that participant’s peak aftereffect. This fitting procedure allowed us to get smooth values that were not constrained to 1 of the 16 target locations. Second, participants’ mean aim direction was calculated as their mean aim during all learning target trials (0° target) in both the learning and probe blocks of the generalization experiment, and participants’ mean movement direction was calculated as their mean hand angle during all learning target trials in the learning and probe blocks. Finally we computed three separate “difference scores” for each participant, corresponding to three

potential centers of generalization: We calculated the difference between the location of their peak aftereffects and 1) the learning target direction of 0° , 2) their mean aiming direction during 0° target trials, and 3) their mean movement direction during 0° target trials (Figure 3.8B).

We performed three one-sample t-tests on the difference scores (Bonferroni corrected) and, as predicted, found no significant difference between the location of participants' peak generalization and their mean aiming directions (Figure 3.8B; $p = 0.35$) but found significant values for both the target location ($p < 0.001$) and mean movement direction ($p < 0.01$). These results support previous findings showing that the selected movement plan may determine the locus of generalization (Day et al. 2016). Furthermore, our results show that individual differences in generalization may be partly explained by participant-specific action selection behavior during a visuomotor learning task.

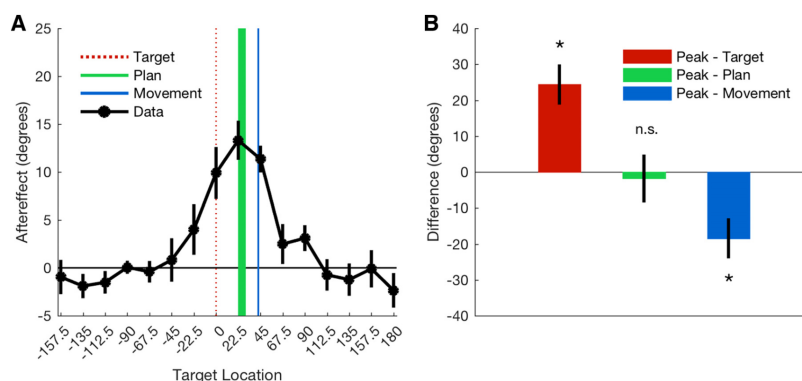


Figure 3.8: The first moment of the implicit generalization function. (A) Generalization of aftereffects. Vertical lines show the predicted center of generalization for target-based generalization (red), plan-based generalization (green), and movement-based generalization (blue). The horizontal width of the green and blue lines represents 1 s.e.m. of, respectively, the mean aim direction and hand angle during learning trials. (B) Difference score analysis. Gaussian functions were fit to each participants' distribution of aftereffects to determine the peak of generalization for each participant. Difference scores were computed by subtracting the learning target location (0°), the participant's mean aim/plan, or the participant's mean movement values during learning trials from their peak generalization value. Error bars = 1 s.e.m. * = $p < 0.01$, Bonferroni-corrected value. n.s. = not significant.

3.2.3 An updated state-space model

Generalization of adaptation can be explained by a population code, where neurons (“units”) with distinct tuning functions are sensitive to a range of movement directions (Donchin et al. 2003; Thoroughman and Taylor 2005; Taylor et al. 2013). When error-based learning occurs at one reach direction, it spreads over to adjacent reach directions. Thus, adaptation should be represented as a vector of states corresponding to multiple directions, rather than a single one. Generalization was incorporated into the 2-state model as follows: The state of implicit adaptation, X^I , was modeled as a 1 x 360 vector $\vec{X}^I$, corresponding to a tiled space of independent “states” at each integer angular location along the circumference of the two-dimensional circular workspace. Surrounding states are updated via a Gaussian function, where the visited state on trial t was set as the center, μ_t , of the Gaussian with standard deviation σ and height B^I , the latter being equivalent to the learning rate of the slow process,

$$\vec{X}_{t+1}^I = A^I \vec{X}_t^I + B^I e^{(\vec{X}^I - \mu_t)^2 / 2\sigma^2} * \varepsilon_t \quad (3.6)$$

The value used to determine the center of the Gaussian on each trial (μ_t) constitutes the frame of reference for generalization; for instance, the target location, or movement plan, on trial t . In the traditional target-based conception of generalization, the center of generalization on every trial (μ_t) is the target location; in our updated plan-based model, the center of generalization is the current aiming direction (Day et al. 2016). Thus, the predicted motor output on trial t (X_t) is the sum of the current explicit state (X_t^E) and the implicit state (X_t^I) associated with the current model’s “center” of generalization, be it the target (0° state) or the plan (X_t^E state, which models the current aiming direction). We used this updated model, and the width parameter gleaned from the generalization experiment (Figure 3.7), to simulate and fit the time course of adaptation in three different “rebound” paradigms (recall Figure 3.3), which are designed to help isolate the fast and slow components

of learning.

After deriving a width parameter for generalization (Figure 3.7C) and confirming that generalization of adaptation is likely determined by the movement plan (Figure 3.8), we simulated the generalization model (Equation 3.6) over three rebound paradigms similar to the one shown in Figure 2.2. The three paradigms differed only in the magnitude of the second rotation — critically, the observed degree of implicit learning in the second rotation block is determined by the change in participant’s explicit aiming plan brought about by the second rotation (Figure 3.9). Thus, the sign of the second rotation shouldn’t matter; a paradigm with a 45° rotation followed by a -45° rotation should yield similar implicit learning to a paradigm with a 45° rotation followed by a 135° perturbation ($\Delta 90^\circ$). Both paradigms will cause a similarly large change in participants’ explicit movement plans, so that uncharted implicit states are visited, driving the observed implicit learning to 0 (Figure 3.9A, B, D, E). However, a modest shift in the second rotation, going from 45° to 75° , should only lead to a subtle dip in observed implicit learning, as the participant’s plan does not shift very much (Figure 3.9C, F). Simulations of the model, showing the time course of explicit and implicit learning, as well as heat maps showing updating in the full space of states/units of the implicit internal model, are shown in Figure 3.9.

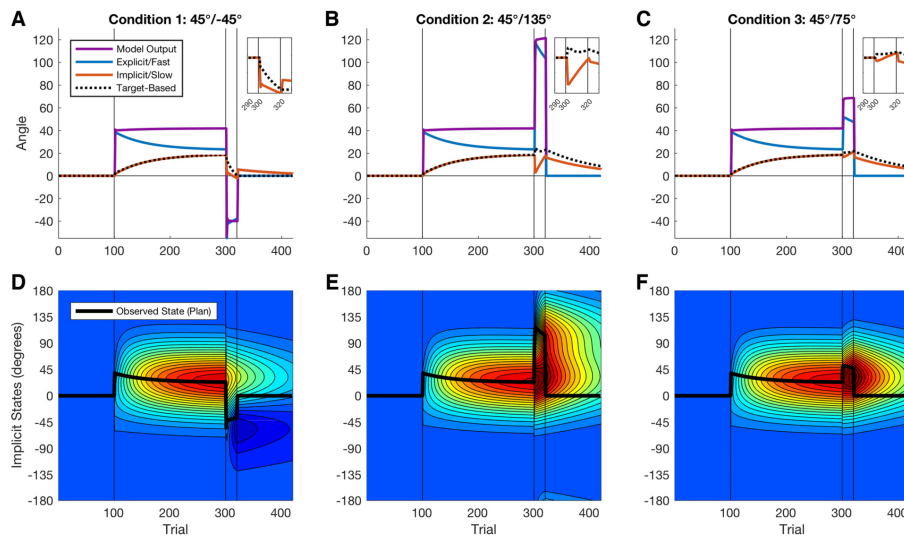


Figure 3.9: Model simulations. (A) Rebound condition 1, model simulation. The plan-based model predicts an implicit adaptation time course (red) that drops quickly at the onset of the second rotation and shows an intermediate value of rebound. A target-based model predicts an implicit time course (dotted line) that shows a weaker drop and no rebound. Insets: zoomed-in implicit simulation data from the end of the first rotation (trials 290–300), through the counter-rotation (301–320), and the beginning of the clamp (321–330). (B) Rebound condition 2. This schedule is identical to condition 1, except that the second rotation is 135°. The plan-based model predicts an implicit adaptation time course (red) that also drops quickly at the onset of the second rotation (like condition 1), then regains ground, and finally shows robust rebound. Critically, a target-based model predicts an implicit time course (dotted line, inset) that shows an increase instead of a drop upon the second rotation. (C) Rebound condition 3. This schedule is identical to condition 1, except that the second rotation is 75°. The plan-based model predicts an implicit adaptation time course (red) that drops only subtly at the onset of the second rotation but shows robust rebound. Importantly, a target-based model predicts an implicit time course (dotted line, inset) that shows a slight increase instead of a drop upon the second rotation. Both models make similar predictions due to the more modest change in the aim direction (blue). (D–F) Heat maps showing learning at each of the 360 “states” during the 3 conditions, using the plan-based generalization model. Hotter colors denote positive adaptation at those states. Black lines show the state being observed/maximally updated during a given trial (in our model, this is the fast process).

The simulations predicted the behavioral results (Figure 3.10), and the model fits were generally strong. Indeed, the plan-based model easily surpassed other generalization models (ΔBIC of plan-based generalization versus movement- and target-based < -50). Figure 3.10 shows the results of the behavioral experiments and model fit of the generalization model (Equation 3.6). Note that panel A shows an experiment with the same protocol as section

3.1, and the discontinuities seen in the second rotation phase and the error clamp phase have been captured by the updated model. The model predicts those discontinuities because if the plan shifts (because of a change in the rotation/environment), a new set of units will be singled out for adaptation. A similar shift occurs in the condition where the rotation increases by 90° (Figure 3.10B), and only a small drop in implicit is seen when the rotation changes by only 30° (Figure 3.10C). All of these results are explained by a model whereby the explicit action plan determines which representations get implicitly adapted in the face of sensory prediction errors.

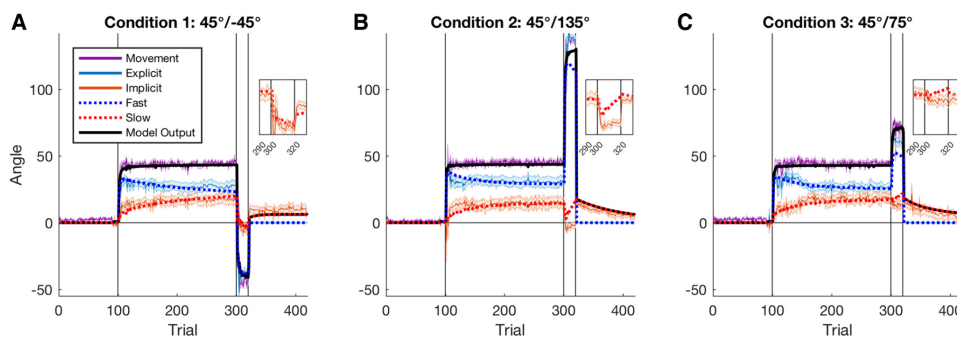


Figure 3.10: Model fits and behavioral results. (A) Data and plan-based model fit for rebound condition 1. (B) Data and model fit for rebound condition 2. (C) Data and model fit for rebound condition 3. Insets: zoomed-in implicit learning data from the end of the first rotation (trials 290–300), through the counter-rotation (301–320), and the beginning of the clamp (321–330). Shading = 1 s.e.m.

In sensorimotor adaptation research, the generalization of adaptation is presented as a spatial phenomenon — generalization reflects the degree of adaptation at naive locations in space relative to a reference point at an over-trained location. This characterization of generalization is modeled in physical space, which is associated with directional or velocity tuned neurons in motor cortex (Paz et al., 2003; but see Churchland et al., 2012) and the cerebellum (Coltz et al., 1999). However, the generalization of explicit planning may not be a question of representing physical space. The shape of generalization functions from various domains of psychophysics — for instance, in auditory and visual systems — are strikingly similar to each other, suggesting a universal law of generalization across an

abstract psychological dimension (Shepard, 1987). It may be that while the Gaussian-like generalization of adaptation (Figure 3.8A) is constrained by a spatially-tuned population code, the factor that seems to dictate the first moment of that function, the explicit plan, is a context-sensitive, abstract psychological construct concerning the dynamics of the learning environment and the goals of the learner. This apparent relationship between high-level and low-level representations in sensorimotor learning provides an intriguing example of multi-system interaction in the CNS, and gives volitional planning a central role.

3.3 General discussion

The studies outlined above reveal several principles of multi-system interaction in the learning of simple sensorimotor transformations. Section 3.1 showed that explicit selection and implicit execution processes contribute to sensorimotor adaptation, and that learning can be modeled by a 2-state model where the explicit process is fast and flexible and the implicit process is slow and robust.

Section 3.2 showed that the representation of implicit learning is influenced by the explicit motor plan, where the plan dictates which units in the implicit state space are updated after errors. This finding is implemented in a state space model with “plan-based” generalization of implicit adaptation. Beyond the implications for generalization itself, these results provide an important framework for thinking about the interaction between selection and execution throughout motor learning. In general, motor skill is equated with the precision of movements, which is certainly partly related to implicit adaptation and cerebellar function, as well as the reinforcement of certain motor commands (Huang et al., 2011). However, if the motor commands to be improved upon are themselves selected by a cognitive system, it stands to reason that the selection system is a fundamental aspect of skill (Stanley and Krakauer, 2013). Consider, for instance, an expert archer competing on a windy day – precision must be maintained not only for movement plans in minimal wind conditions, but for

the full set of actions that could be selected across a range of wind conditions.

The plan-based model of section 3.2 (Equation 3.6) has specific neural implications, suggesting that explicit action selection serves as the input to the internal representation that is modified by sensory prediction error-based adaptation (McDougle et al., 2017). This interpretation was echoed in a recent study (Sheahan et al., 2016): Using an interference paradigm, where two opposing force field perturbations are induced on the same movement direction (on different trials), Sheahan and colleagues showed that interfering perturbations can be learned if each is associated with a different movement plan, operationalized by different “follow-through” targets that participants would occasionally have to move toward at the end of the reach (i.e., after pushing through the force field). Without allowing for distinct plans, massive interference (and thus, no learning) is usually observed in such tasks. Critically, disparate plans can separate motor memories even if the full follow-throughs are not actually performed, suggesting that the neural states produced by the planning stage of movement — but not the actual execution of those plans — are the critical factor determining error-based adaptation (Sheahan et al., 2016; Hirashima and Nozaki, 2012).

These results suggest that the planning stage determines which internal representation is then updated by subsequent sensory prediction errors (McDougle et al., 2017). The neural interpretation is that the plan sets the initial conditions of a neural trajectory in the downstream dynamical system that directly controls motor output (Sheahan et al., 2016; Churchland et al., 2012). Consistent with this interpretation, recent results using a brain machine interface (BMI) in nonhuman primates revealed that training on a visuomotor rotation “covertly” (i.e., using the BMI but not eliciting actual arm movements) transfers to overt motor behavior on that same rotation, and that this transfer may be due to corresponding motor cortical activity in the planning stages of both contexts (Vyas et al., 2018). It is likely that in most settings, this planning stage involves explicit selection of a desired action.

Taken together, the results in this chapter show that multiple interacting processes with different time courses drive the acquisition of a modified sensorimotor map, with a fast, cognitive selection process quickly closing large errors and setting the stage for the slow, implicit calibration of an internal model. In the next chapter I investigate the computations driving these cognitive selection processes.

4 Action selection strategies: working memory and cognitive motor learning

Learning a new motor skill — say, how to shoot an arrow — can be aided by understanding the causal structure of the environment, for instance, that a wind from the east will tend to push the arrow west. This kind of explicit knowledge can shape how an agent selects their movements (e.g., aiming at a slight angle into the wind). Although it is clear that such heuristics have a role in sensorimotor learning (and are ubiquitous in the teaching of skills to others), details concerning the psychological representations underlying these heuristics, how they may be applied during learning, and their associated neural substrates, are not well understood.

Even in a much simpler motor task, like the visuomotor rotation task (section 3; Krakauer, 2009), participants appear to leverage strategies to rapidly improve performance, effectively “bootstrapping” the learning curve (Bond and Taylor, 2015; McDougale et al., 2017). These strategic processes appear to be related to higher reaction times (Haith et al., 2015), improved task performance (Fernandez-Ruiz et al., 2011), and the direction of eye gaze (Rand and Rentsch, 2015; de Brouwer et al., 2018). Increased reaction time, and the fact that strategies are often verbalizable, suggests that they may reflect a form of deliberative, controlled processing rather than automatic processing (Cohen et al., 1990). Controlled processes appear to be reliant on working memory, making it one candidate system that may underlie the observed strategies in motor learning. Indeed, evidence suggests that spatial working memory ability correlates with performance in visuomotor learning (Seidler et al., 2012), and recruits similar neural circuits (Anguera et al., 2010). Moreover, spatial working memory ability specifically correlates with the use of explicit strategies in visuomotor rotation tasks (Christou et al., 2016). However, it remains unclear what operations of working memory directly drive strategies for motor learning.

We hypothesized that explicit action selection takes two broad forms: 1) learning and applying a parametric rule governing the desired sensorimotor transformation (e.g. mentally transforming a movement goal), and 2) acquiring and storing discrete stimulus-specific responses (i.e. associating specific stimuli with specific movements). In this chapter I provide evidence for these two distinct behavioral policies, both of which are established functions of spatial working memory (Georgopoulos and Pellizzer, 1995).

Consider someone first learning to play the piano. Ideally, she would instantly be able to see the notes on the staff and immediately know the appropriate piano keys to play. However, as anyone who has ever attempted to play a new instrument knows, reading music is simple when the musical score only has a few notes in a small range, but as it gets more complicated things fall apart — if we just consider the number of piano keys in an octave, it easily exceeds our working memory capacity (Miller, 1956). A common strategy to overcome this limitation is to approach the task “parametrically”: She can anchor her thumb on middle C and reference other notes on the lines of the musical staff relative to this piano key. Indeed, this is a common strategy used in beginner piano pedagogy. While this strategy affords her the ability to play a more complex melody within a few minutes of practice, it becomes increasingly cumbersome the further a given note is from middle C — bearing remarkable similarity to the phenomena of mental scanning and mental rotation (Kosslyn et al., 1978; Shepard and Metzler, 1971). These two strategies, one discrete (caching each response-note relationship) and one parametric (computing the distance from C), offer two ways to approach learning a complex novel motor skill.

Mental rotation (MR) is one canonical example of a parametric, or “analog” computation (Shepard and Metzler, 1971; Cooper, 1976), in which an internal mental representation is manipulated like an external physical object. Visual working memory has been highlighted as the substrate for classic MR (Hyun and Luck, 2007). In the motor domain, evidence from behavior and neurophysiology provides support for MR in the planning of reaching

movements: Reaction time (RT) during the mental rotation of a reach goal is a linear function of the magnitude of required rotation, sharing remarkable similarities to the mental rotation of visual percepts (Georgopoulos and Massey, 1987; Pellizzer and Georgopoulos, 1993; Bhat and Sanes, 1998). Moreover, the decoded direction of neuronal population vectors in motor cortex appears to “rotate” through direction space during the mental rotation of a reach goal (Georgopoulos et al., 1989; but see Cisek and Scott, 1999).

Critically, MR is sensitive to an arbitrary number of signs, magnitudes, and even planes of rotation (Shepard and Metzler, 1971), and thus constitutes a flexible computation that appears to rely on a straightforward heuristic, one that is formally equivalent to applying an n -dimensional rotation matrix with angle θ to a representation in working memory, in this case, the vector of a reach relative to a target:

$$\begin{bmatrix} reach_x \\ reach_y \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} target_x \\ target_y \end{bmatrix}$$

Similar heuristics appear to exist for other sensorimotor transformations — for instance, MR-like RT effects are observed when humans compute gain perturbations on reach goals (Bhat and Sanes, 1998). Here, the heuristic would simply represent a scalar transformation.

Response caching (RC), on the other hand, is here defined as the maintenance of acquired associations between a set of stimuli and a set of responses maintained in working memory (Collins and Frank, 2012), relying on processing in the prefrontal cortex (Petrides 1985; Curtis and D’Esposito, 2003). As a form of discrete item-based working memory, the efficacy of a discrete RC strategy should be subject to both load (e.g. the number of items to be stored) and interference (e.g. the introduction of intervening items between trials). Discrete RC should also be computationally inexpensive in simple learning contexts, but grow in cost and decline in accuracy as the learning environment becomes more complex (Cowan, 2010; Collins and Frank, 2012).

Stimulus	Response
$target(\theta_1)$	$reach(\theta_2)$
$target(\theta_3)$	$reach(\theta_4)$

MR, on the other hand, putatively operates on a single generalizable rule, and should thus not be particularly vulnerable to load and decay effects.

In section 4.1 I revisit an older data set to show that RT is a reliable measure of explicit learning. In section 4.2 I provide evidence for two signatures of working memory processes in VMR data. In section 4.3 I provide evidence that both RC and MR are recruited in VMR, where MR appears to be recruited early in learning while the subject infers and implements the structure of the task (i.e. the particular rotation direction and magnitude), after which RC is recruited given that the load (i.e. number of stimuli) is sufficiently low. In section 4.4 I further elucidate the computations underlying these two policies, showing how time pressure during the RT reveals the putative dynamics of MR and RC within a trial. Lastly, in section 4.5 I present the results of two neuropsychology experiments, investigating the role of the cerebellum and medial temporal lobe in explicit motor learning. Taken together, these results reveal new details about the action selection processes of motor learning, with implications for structural learning, generalization, and the role of explicit knowledge in skill formation.

4.1 The validity of reaction time data as a reflection of selection processes in sensorimotor adaptation

Mental chronometry – the careful study of reaction time (RT) – has been a useful tool in cognitive psychology for many years, dating as far back as Donders (1868). Most relevant to the current study, RT has proved a vital tool for the study of controlled vs. automatic processes (Stroop, 1935). Because RT is a central component of this chapter, the first step

was to investigate which motor learning process (explicit vs. implicit) is associated with RT in the VMR task.

Data from 55 participants performing an aim-report VMR task similar to that described in chapter 3 were analyzed (Taylor et al., 2014). In this experiment, participants reached to 8 different target locations, and performed 56 baseline trials, 320 rotation trials, and 80 washout (aftereffect) trials.

The data were analyzed as follows: First, because of linear trends in both the RT and learning time courses during the task, all three input variables (RT, explicit learning, and implicit learning) were detrended using linear regression for each subject (Figure 4.1 top). After detrending, a general linear model was fit to measure the effects of explicit and implicit learning on trial-by-trial RT.

As shown in Figure 4.1 (bottom), explicit learning (i.e. the detrended magnitude of the angle reported by the subject on each trial) positively predicted RT ($p < 0.001$). Implicit learning (i.e. the detrended magnitude of the difference in angle between movements and reported aim directions on each trial), however, did not significantly predict RT ($p = 0.12$). Lastly, the regression weight for explicit learning was significantly greater than for implicit learning ($p = 0.002$).

This pilot analysis highlights two important things: 1) RT in VMR is modulated by explicit but not implicit learning, and 2) the actual magnitude, in degrees, of explicit learning appears to linearly scale RT (i.e. the further the selected plan was from the target, the longer the RT). The latter point will be returned to in the following chapters.

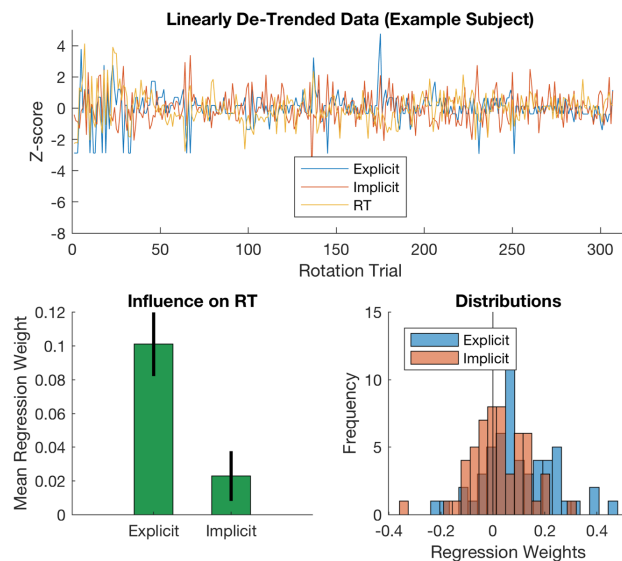


Figure 4.1: Effect of explicit and implicit learning on RT. Top: de-trended representative subject’s data for explicit learning (blue), implicit learning (red), and reaction time (yellow). Bottom: results of regression analysis. Error bars = 1 s.e.m.

4.2 Evidence for working memory signatures in explicit selection processes

As described earlier, strategic learning processes occur alongside implicit ones (Taylor et al., 2014; Huberdeau et al., 2015), and these processes appear to be related to both higher reaction times and better performance (Fernandez-Ruiz et al., 2011; Haith et al., 2015). The explicit “reportability” and increased reaction time related to strategic learning imply the presence of “controlled” processing (Cohen et al., 1990).

In light of these findings, we hypothesized that classic working memory effects could be observed in VMR tasks, with response caching (RC) reflecting discrete working memory computations that are subject to temporal decay, and mental rotation (MR) reflecting a parametric computation that is affected by rotation magnitude. We also predicted a trade-off between working memory processes depending on set size (i.e. number of targets), though this is not explicitly tested until section 4.3.

First, previous VMR data were re-analyzed to look for a general effect of “set size” on RT, which is one common signature of working memory (Sternberg, 1965). The task was a classic VMR reaching task (Figure 3.1), with a fixed 45° rotation on every rotation trial (data are from Bond and Taylor, 2015), and verbal aim reporting. Critically, the number of target locations, and thus, the number of movement solutions, varied across groups (four groups: 1, 2, 4, and 8 targets, respectively). As predicted by discrete, capacity-limited working memory, the number of targets participants were exposed to during learning correlated with increased RTs (Figure 4.2A), and this held true even at the end of the learning block (main effect of number of targets: $p < 0.05$ for last 20 trials). Moreover, equivalent RTs across groups in the baseline period suggest that the RT results were not merely a function of visual search time differences between groups.

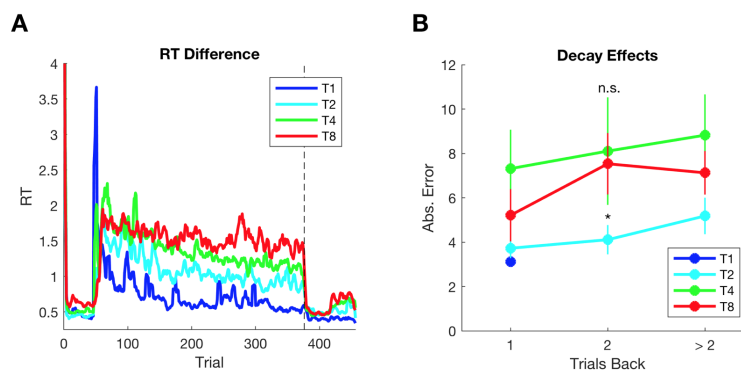


Figure 4.2: Effect of number of targets on (A) RT and (B) movement error after various intervals between successive target presentations. Error bars = 1 s.e.m.

Figure 4.2B reveals another hallmark of working memory – decay (Peterson and Peterson 1959; Collins et al., 2012). That is, movement errors were larger if the interval (i.e. number of intervening trials) between presentations of the same target location was longer, suggesting time-dependent decay in a working memory trace and/or interference. Critically, this effect was only significant in the 2-target group ($p < 0.05$), but not in the 4- and 8-target groups. This suggests that, perhaps, the 4- and 8-target groups were not relying on a simple “look-up table” strategy – which would be subject to decay – but could be instead adopting a rule-

based strategy. Indeed, human working memory capacity for accurate S-R relationships is generally somewhere between 3-4 items (Cowan, 2010), consistent with the hypothesis that the 4- and 8-target groups are using a different strategy.

One such candidate for a rule-based strategy is a parametric manipulation in working memory, like mental rotation (MR; Shepard and Metzler, 1971). In MR, the degree of rotation – of a visual percept (Cooper, 1975) or a movement plan (Georgopoulos and Massey, 1987) – that a person needs to carry out linearly scales their RT. Preliminary evidence for such a process in VMR tasks should be clear – a larger rotation should produce a longer RT. I thus re-analyzed another data set (from Bond and Taylor, 2017), where the rotation magnitude varied across trials. Rotation sizes were uniformly randomly varied in “mini-blocks” of 8 trials each (Figure 4.3A). After converting rotations to absolute value (to get raw magnitude), I binned them in 15° bins. The first trials of each bin were excluded, as participants had not seen the new rotation yet (leaving 305 rotation trials total).

As shown in Figure 4.3B, there was a clear effect of rotation magnitude on RT: Higher rotation magnitudes led to longer RTs (t-test on regression weights: $p < 0.001$). This result is consistent with the mental rotation of a reach vector, as proposed by Georgopoulos and colleagues (1989; 1993). Taken together, Figures 4.2 and 4.3 provide compelling preliminary evidence for at least two putative working memory processes in explicit VMR learning – a *discrete* “look-up table” of movement goals that is subject to decay and is active in low set sizes, and a *parametric* “rule-based” computation that implements a generalizable analog computation that is active in high set sizes. This hypothesis was tested directly in the ensuing sections.

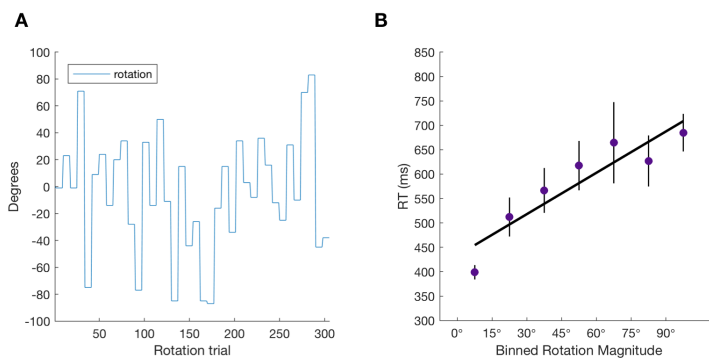


Figure 4.3: Evidence for mental rotation of a movement plan. (A) Example rotation schedule for one participant. (B) Effect of rotation magnitude on reaction time. Error bars = 1 s.e.m.

4.3 Two dissociable working memory strategies in motor learning

Here we further tested the hypothesis that two broad working memory strategies could describe explicit learning in VMR – one leveraging discrete representations and one leveraging parametric representations. Moreover, we reasoned that participants would initially use a parametric strategy to discover the structure of the task, and, in the case of a low set size, transition to a discrete strategy over time. Returning to our piano example, this would involve an eventual abandoning of the parametric “middle C” strategy, to be replaced by a rich stimulus-response map.

As evidenced by previous data sets, strategic learning appears to assume at least two broad forms — discrete response caching (RC) of stimulus-response contingencies, and the implementation of time-consuming parametric computations (e.g. mental rotation, MR). Here we show that participants’ reaction times and patterns of movement errors suggest that both action selection policies are employed over the course of learning, and appear to trade off based on both the progress of learning and the complexity of the learning environment (i.e. set size).

4.3.1 Task design and method

80 right-handed participants (age range 18-24) were recruited from the research participation pool maintained by the university in exchange for course credit or monetary compensation. Handedness was verified using the Edinburgh handedness inventory (Oldfield, 1971). All participants participated in accordance with the Princeton's Institutional Review Board and provided written, informed consent. A sample of 20 participants was used per condition. In all experiments, participants made rapid, center-out reaching movements to visual targets (5.0 mm radius) using the digitizing tablet described in previous chapters. The task was controlled by custom software written in MATLAB (Mathworks, Natick, MA; Psychophysics Toolbox). A small cursor (2.5 mm radius) provided endpoint feedback after each reach terminated. Analyses were conducted in MATLAB and R (GNU).

The experiment was a blocked-design reaching task using a 2 X 2 between-participants design, crossing the factors rotation magnitude and number of targets ("set size"; Figure 4.4). For the rotation factor, two magnitudes were used, $\pm 25^\circ$ and $\pm 75^\circ$, with the sign of the rotation counterbalanced within each condition. For set size, participants were exposed to either two targets (2T) or 12 targets (12T). In the 12T condition, target locations were pseudorandomly presented at 12 possible locations (0° , 30° , 60° , 90° , 120° , 150° , 180° , 210° , 240° , 270° , 300° , 330°), and the same location was never repeated in two consecutive trials (Figure 4.4, bottom right). In the 2T condition, two targets were presented, and the locations were randomized across participants to include one position from the 12 possible locations above and its opposite 180° away (Figure 4.4, top right). Within a participant, targets were presented pseudorandomly, where the same target was not repeated more than twice in a row.

In a baseline block (36 trials), participants reached to the targets with veridical endpoint feedback. At the start of the subsequent rotation block, the rotation was abruptly applied and was maintained for 300 trials, all with endpoint feedback. Finally, in the washout block,

participants were told to cease any strategy they may have adopted to counter the rotation and reach directly for the targets. Critically, to limit implicit sensorimotor adaptation and better isolate strategic explicit learning, all feedback was delayed by 2 s on every trial (Kitazawa et al., 1995; Brudner et al., 2016). Delaying spatial feedback has been shown to abolish implicit motor adaptation and accompanying aftereffects, and this result has been linked to temporal constraints in cerebellar associative learning (Kitazawa et al., 1995). This delay manipulation was crucial, allowing us to analyze learning curves free of contamination by implicit adaptation.

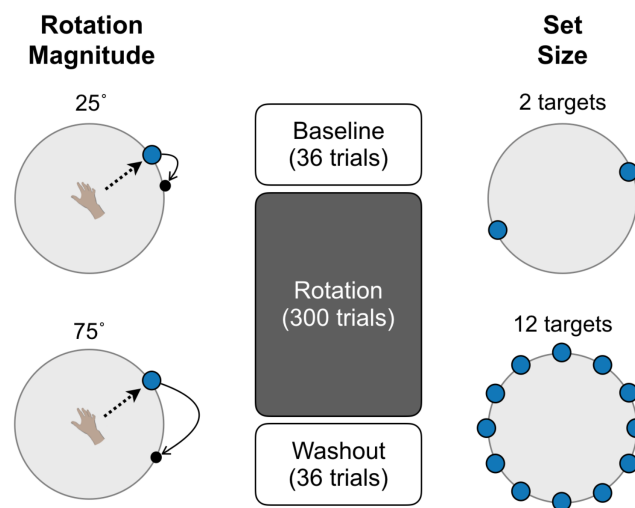


Figure 4.4: Task design. Participants performed a standard visuomotor rotation task, attempting to land a rotated visual cursor on a target. A between-subjects, 2 X 2 design was used, crossing the factors Rotation Magnitude (the size of the rotation in the rotation block; left) and Set Size (the number of possible target locations in the task; right).

Movement angles were computed as the angle of the hand, relative to the target, when it crossed the invisible target ring. All movement angles were rotated to a common reference frame about the 0° axis of the unit circle. In all experiments, reaction time (RT) was measured as the time elapsed from target appearance to the hand leaving the start circle (i.e. 5mm from the center of the tablet), and movement time (MT) was measured as the duration of the movement from the end of the RT to the time at which the hand crossed the invisible workspace ring. In analyses with normalized absolute errors, angular errors

were scaled between 0 and 1 for each participant. Lastly, trials where RTs exceeded three standard deviations above or below a participant's mean RT over the entirety of the task were excluded from analyses (1.6% of trials).

4.3.2 Results

As shown in Figure 4.5A, participants' RT time courses (averaged over 12-trial cycles) appeared to be affected by the conditions, with both target number and rotation magnitude showing effects. Moreover, participants' movement angle time courses (also binned) show successful countering of the rotation in all groups (Figure 4.5B).

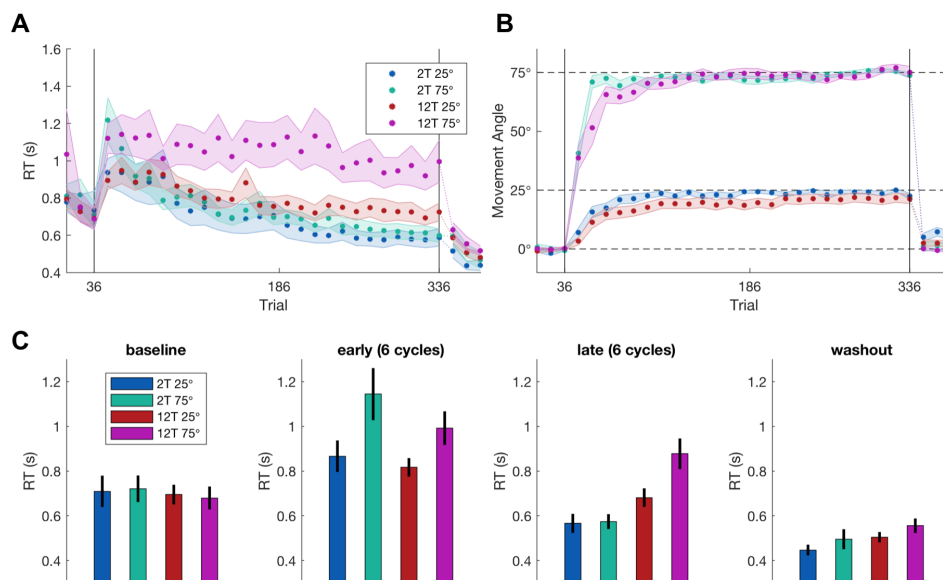


Figure 4.5: Results. (A) Time course of mean RTs, averaged over 12-trial bins. Vertical lines delineate the task block (baseline, rotation, and washout). (B) Movement angle across the groups, averaged over 12-trial bins. (C) Mean of median RTs averaged over the cycles of trials in the beginning (early) and end (late) of learning, and the full baseline and washout blocks. Shading and error bars = 1 s.e.m.

To perform statistical comparisons, we averaged trials by cycles that controlled for differences in the number of exposures of each target location (Figure 4.5C). Thus, a cycle was 2 trials long in the 2T group and 12 trials long in the 12T group. Bins of 6 cycles were com-

pared between early learning (first 6 cycles) and late learning (last 6 cycles). Two separate 2-way ANOVAs were used, one early and one late, to test for effects on RT and movement angle by set size (2T vs. 12T), rotation magnitude (25° vs. 75°), and their interaction.

Figure 4.5C shows the results of this analysis on RT. First, we found that in the baseline block, there were no significant effects of set size or rotation magnitude (all p 's > 0.60). Thus, the number of possible targets did not itself modulate RT when no rotation was present.

Comparing early and late learning, we found a significant main effect of Time ($p < 0.001$), reflecting the fact that all groups decreased their RT over training. A significant effect of Rotation Magnitude was also observed, suggesting that participants may be employing a parametric MR strategy — larger rotations resulted in larger RTs ($p = 0.005$). While we did not find a main effect of Set Size ($p = 0.34$), we did find a significant two-way Time X Set Size interaction ($p < 0.001$) and, more importantly, a three-way Time X Set Size X Rotation interaction ($p = 0.036$). Two post hoc t-tests (Bonferroni-corrected) revealed an effect of Rotation Magnitude on late learning RT only in the 12T group ($p = 0.019$), but not in the 2T group ($p = 0.88$). Taken together, these results are consistent with the hypothesis that all groups may have used a parametric MR strategy to find the task solution early in learning, and then either stayed with MR for the remainder of the task (12T group), or transitioned to a discrete RC strategy, where RT was not modulated by rotation size (2T group).

This transition can be seen in Figure 4.6, where I've have plotted the difference between mean RT time courses of the two rotation magnitudes for the 2T and 12T groups, respectively. The prediction line is the difference in RT predicted by an independent experiment on mental rotation in VMR (see next section). Both groups appear to perform MR early in learning, but the 2T group appears to drop that strategy after several trials (this analysis will be revisited later).

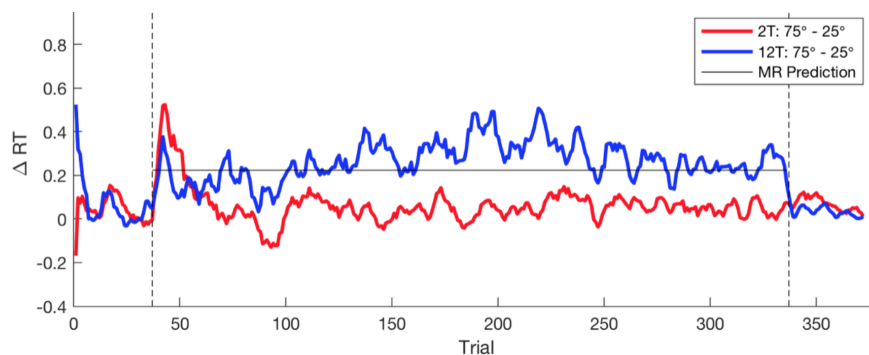


Figure 4.6: RT effects by learning progress. The red line depicts RT differences between the two 2-target groups, and the blue line depicts RT differences between the two 12-target groups.

Lastly, no significant effects of rotation magnitude or set size were found for early or late movement angle, though trends are visible (Figure 4.5B; all p 's > 0.06).

Two secondary analyses also comport with our model of discrete versus parametric working memory strategies. The first analysis was a *post hoc* analysis: We looked at the participants' "sign" errors (i.e. trials where subjects reached in the wrong clockwise/counter-clockwise direction). Normatively, discrete RC and parametric MR make different predictions about sign errors: In RC, sign errors should represent trials where participants accidentally extract the wrong discrete response held in working memory ("swap" errors; Bays, 2016). This would constitute a -155° error in the 25° conditions, and a -105° error in the 75° conditions. In MR, however, sign errors should represent trials where participants accidentally "flip" the rotation angle, thus reaching to the correct magnitude but in the wrong direction. This would constitute a -25° error in the 25° conditions, and a -75° error in the 75° conditions.

As sign errors were relatively rare, we pooled sign errors in each group. Sign errors were designated as such if the participant reached $< -15^\circ$ from the target. As displayed in the histograms in Figure 4.7, swap errors were the more common sign error in the 2T groups, consistent with a discrete, item-based working memory representation (Bays, 2016). Flip errors, however, were the more common error in the 12T groups, consistent with a

parametric representation that occasionally computes an incorrect sign of rotation.

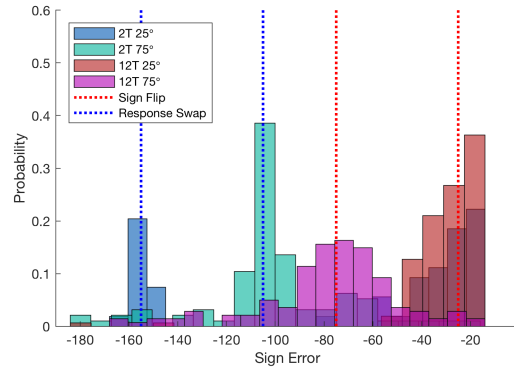


Figure 4.7: Swap errors vs. flip errors. Histograms of pooled “sign error” data in each group. Vertical dotted lines represent predicted locations of “swap” errors (generating the wrong response given two choices) and “flip” errors (reaching relative to the correct magnitude of rotation but the wrong sign).

Lastly, in Figure 4.8 we replicated our decay analysis from the previous section (Figure 4.2B). We hypothesized that in the case of a parametric computation (12T group), decay would not be expected as the rotation rule is the same in every trial. To assess the effect of decay on reach error, we first combined the groups based on set size, and then normalized each participant’s absolute reach errors during the rotation block (to account for error differences across rotation magnitudes). We then classified errors based on the number of trials elapsed since the current target was previously encountered (range: 1-3 trials for 2T groups, 2-23 trials for 12T groups). The interval was cut off at 3 in the 12T group to allow for ease of visualization. As predicted, we found a significant effect of decay in the 2T groups (t-test on slopes: $p = 0.002$), but not in the 12T groups ($p = 0.74$), consistent with decay and/or interference of a discrete working memory trace (Collins et al., 2012).

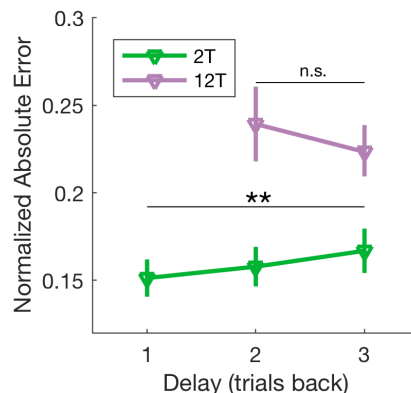


Figure 4.8: Decay of response accuracy. Normalized absolute reach errors, grouped by the interval between the current trial and the previous trial in which the same target was presented. Error bars = 1 s.e.m. ** = $p < 0.005$

4.4 Limiting reaction time to reveal dynamics of action selection

Having established our model of parametric mental rotation (MR) and discrete response caching (RC) comprising two cognitive action selection strategies in visuomotor transformations, we wanted to further confirm that these mechanisms were valid models of actual within-trial mental computations. To do this, we decided to adopt a “forced reaction time” task (Ghez et al., 1997), which limits the amount of time participants have to respond in the task. This procedure has recently been used to show how different learning processes progress in VMR (Haith et al., 2015). Here, however, we wanted to test how rapidly-induced movements could reflect the state of cognitive processing (Shepard and Metzler, 1971; Cooper, 1976).

Our specific hypotheses were as follows: Interrupting MR should induce intermediate movements that represent partially-rotated movement plans, but interrupting RC should induce bimodal movement distributions with modes at each cached movement location, but no significant intermediate movements. Such a result would provide evidence that participants do indeed leverage different working memory representations depending on the specific task context (Georgopoulos and Pellizzer, 1995).

4.4.1 Task design and method

To investigate MR computations, we used two tasks: One task (FREE task) had no constraints on RT – this was essentially a control task designed to replicate our previous findings that, during MR of a reaching movement, RT is a linear function of rotation magnitude (Figure 4.3). The second task (FORCED task) used the forced-RT procedure mentioned above, and was designed to reveal the dynamics of MR in action.

Subjects ($N = 32$) performed both tasks (FREE and FORCED; Figure 4.9), with the order of completion counterbalanced, as well as the rotation sign used in the FORCED task (-90° versus 90°).

FREE Task (Figure 4.9, top): At the start of every trial, participants used continuous cursor feedback to position their hand in a central “start” position (5 mm radius). After holding for 400 ms, a visual target appeared and cursor feedback was removed. Participants were instructed to make a rapid, straight shooting movement to the target. The trial concluded when the hand crossed an invisible ring (8 cm radius), at which point feedback of the cursor was provided at the location where the hand crossed the ring. Slow movement times (MT) were discouraged: If MT exceeded a 700ms limit, a “too slow” warning was delivered visually to the subject by the task software. In line with the instructions, participants moved straight and rapidly, with a mean MT of 208.32 ± 8.54 ms (standard error of the mean).

Participants were instructed that they would be performing a number of trial “pairs”: In the first trial of each pair, the “learning” trial (Figure 4.9, top left), participants were instructed to reach directly at the displayed target and observe and remember where the feedback cursor landed. In learning trials, the target was blue and appeared in one of four off-cardinal locations (10° , 100° , 190° , 280°). In the second trial of the pair, the “execution” trial (Figure 4.9, top right), participants were told to apply what they learned about the relationship between their movement and the resultant feedback and attempt to make the cursor terminate within the target. In execution trials, the target was red and appeared in

one of the three locations in which the learning target did not appear. Target locations were pseudo-randomized within and between trial pairs. Participants performed 140 trial pairs. This task was modeled after a previous study (Pellizzer and Georgopoulos, 1993), but with the distinction that participants were not provided with an explicit symbolic cue regarding the exact solution to the rotation; instead, participants had to determine the rotation's size and sign themselves. Therefore, our task was essentially a single-trial learning version of a canonical visuomotor adaptation paradigm (Krakauer, 2009).

Rotations used in the learning trials ranged from -90° to 90° by 15° intervals. Rotations were uniformly samples throughout the task, and each rotation was seen on 10 different trial pairs, except for the 0° null rotation, which was seen on 20 trial pairs. Each participant received a different schedule of rotation magnitudes and target locations.

Analyses were conducted on execution trials only, limited to trials where there was an imposed rotation. Movement angles were computed as the angle of the hand, relative to the target, when it crossed the invisible target ring. All movement angles were rotated to the same 0° target axis and matched to a single sign for analysis. The first analysis was a regression on participants' reaching angles, relative to the imposed rotation, on execution trials. We used the absolute reaching angle since on a significant number of trials (8.71%), participants approximated the magnitude of the rotation but misinterpreted the sign (i.e. were within 15° of the “flipped” solution; recall Figure 4.7 for an analysis of flip errors). No significant difference in mean RT was found between correct and “flipped” execution trials ($t(31) = -0.99, p = 0.33$), suggesting that participants did not hesitate and change their mind on flipped trials, but simply misremembered the proper clockwise/counterclockwise direction at the onset of the execution trial.

A one-sample t-test was performed on the fitted slopes resulting from the regression of rotation magnitude and movement angle to test for the presence of a reliable trend. The second analysis was a regression of each participant's realized movement angles onto cor-

responding RTs (i.e. mental rotation). A one-sample t-test was performed on the fitted slope values to test for the presence of a reliable trend. Moreover, each participant's slope derived from this regression analysis served as their mental rotation “pace” in the units of milliseconds per degree. We note that we also performed a similar linear regression using the imposed rotation angles as the predictor variable to confirm that it echoed the regression using the actually-realized movement angles; both analyses yielded comparable results, though the movement angle regression is likely to be a better estimate of a participant's true mental rotation pace (Georgopoulos and Massey, 1987).

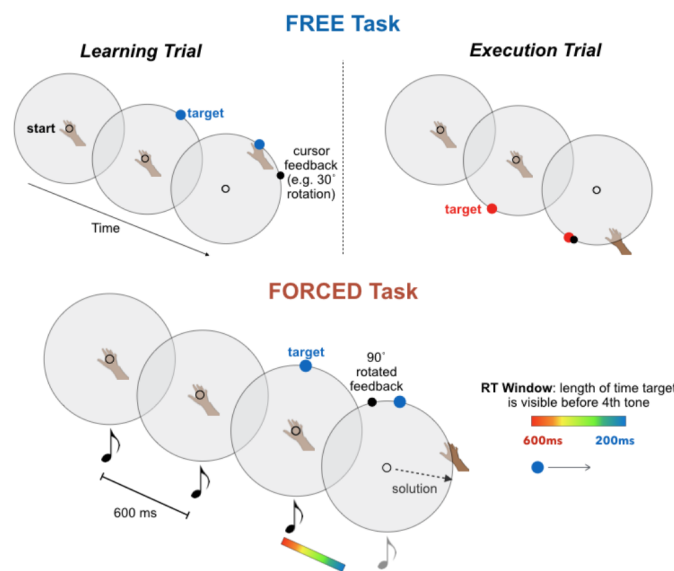


Figure 4.9: Task design. **FREE task:** Participants performed trial pairs consisting of learning trials (left) and execution trials (right). Thirteen rotation sizes were pseudorandomly presented ($-90^\circ:15^\circ:90^\circ$). **FORCED task:** Participants were required to initiate their movement after target presentation and 100 ms after the fourth tone. Targets could appear in either one of 12 locations. The moment of target appearance was titrated to induce participants to react with a controlled distribution of RTs.

FORCED Task: Unlike the **FREE** task, the **FORCED** task utilized a modified forced-reaction-time paradigm to interrupt putative mental rotation (Ghez et al., 1997; Haith et al., 2016; Figure 4.9, bottom). At the start of every trial participants used continuous cursor feedback to position their hand in a central “start” position. Once participants were

positioned in the start, a countdown of four tones was emitted from the computer speakers (Logitech). The tones were played 600 ms apart, and participants were instructed to try and synchronize the initiation of their reach with the fourth tone, effectively “replacing” that tone (Figure 4.9, bottom). The experiment and instructions were deliberately designed to emphasize early movements over late ones: If the participant initiated their movement > 100 ms after the fourth tone, the screen was blanked and a “Too Late” message was displayed. Participants could initiate their movement any time after target appearance, to encourage early RTs. However, if they moved before the target appeared the screen was blanked and a “Too Early” message was displayed. Similar versions of this forced-response-time task have been previously used to study the effect of restricted RTs on visuomotor learning and movement preparation (Ghez et al., 1997; Haith et al., 2016).

Participants were instructed to prioritize reaching on time, and, secondarily, to try and land the cursor on the displayed target. On each trial, the targets could appear in one of 12 off-cardinal locations (10° , 40° , 70° , 100° , 130° , 160° , 190° , 220° , 250° , 280° , 310° , 340°). Endpoint cursor feedback was presented after the participants passed the invisible target ring, and the appearance of the feedback was delayed by 500 ms. Like the previous experiments (section 4.3), this brief delay was added to inhibit implicit adaptation to the imposed rotation and thus isolate the cognitive re-aiming process (Kitazawa et al., 1995; Brudner et al., 2016).

The moment of target appearance was titrated such that participants had varying amounts of time with which to compute the target location, plan, and execute their movement. These windows reflected the time elapsed between target appearance and the fixed moment of the fourth tone. Seventeen test RT windows were used, ranging from 200 - 600 ms by 25 ms intervals. An asymptotic RT window was also used, giving participants up to 1200 ms to react. We reiterate that because participants could move early in our variant of this task (i.e. any time after target appearance), the windows acted as “guides” rather than being

perfectly predictive of participants' realized RTs (i.e. participants tended to move before they needed to). Finally, catch trials of 0 ms were used to ensure that participants stayed on task and executed movements on time (Haith et al., 2016) even if they had not perceived the target yet.

Participants performed the FORCED task in three blocks. To get accustomed to the task, in the first block participants performed 64 baseline trials with veridical endpoint cursor feedback, with pseudo-randomized RT windows and target locations. In the subsequent rotation block, participants performed 624 trials, with 35 trials at each of the test RT windows, 14 trials at the asymptotic RT window, and 15 catch trials. For the rotation block, a fixed rotation of 90° (or -90° , for counterbalancing) was imposed on the cursor. Participants were thoroughly educated about the rotation before this block began. Participants were told to try and counter the rotation and land the cursor on the target every trial. Due to the difficulty of the task, participants were encouraged by a monetary bonus of up to 5 dollars (USD) based on their performance in the rotation block, and were informed that missed trials (reaching too late or too slowly) would count against their performance score. Finally, in the aftereffect block, participants performed 48 trials with a 1200 ms RT window, pseudo-randomized target locations, and no cursor feedback. Participants were told to reach directly for the target on every trial of the aftereffect block. These data were later compared to the baseline data to get an estimate of any implicit adaptation.

Analyses were conducted on trials where participants reached on time (78.90% of trials). Like experiment 1 and the FREE task, movement angles were computed as the angle of the hand relative to the target when it crossed the invisible target ring. Importantly, participants followed the instructions and made straight shooting movements: Movement times were rapid (128.13 ± 6.01 ms), and no feedback was provided during the reach, which helped to ensure movements were straight. Movement speed was computed by taking the average of the derivative of hand position from the time participants left the start circle to the time

they crossed the invisible target ring. (Due to the rapid “shooting” movement required by the task, average speed was used instead of peak speed to reduce noise in the estimate — participants often reached peak speed after passing the target; we note that this particular approximation of speed did not influence conclusions from the speed analyses).

Our first analysis involved investigating participants’ reach angles as a function of RT. RT bins were taken every 25 ms, from 0 ms through 400 ms, with the final bin including all RTs above 400 ms. Similar to previous results (Haith et al., 2016), in catch trials (0 ms RT window) participants prioritized the timing demands of the task when no target appeared and moved relatively randomly around the circle, with some slight biases (data not shown).

A single “critical RT bin” was computed, after which reaches were determined to be primarily non-random (bin 7; $RT > 150$ ms). This bin was determined by conducting serial t-tests on reaching distributions to determine when they displayed a mean that differed significantly from zero. Linear regressions were performed on participants’ full distribution (i.e. no binning) of reach angles and RTs after this “critical” RT bin, and a t-test was performed on the resulting slopes to test for the presence of a significant trend. These slopes were used as the mental rotation “pace” parameters in further analyses.

Mental rotation paces were compared between the FREE and FORCED tasks in four ways: First, a one-sample t-test was performed to show that the null (i.e. that the slopes were statistically the same) could not be rejected. Second, a Bayes Factor was computed on the resulting one-sample t-value using the JZS method (Rouder et al., 2009) to quantify evidence for the null. Third, both parametric (Pearson) and non-parametric (Spearman) correlations were conducted between the values to test for a significant relationship that is robust to outliers (both metrics were used for all correlations in this study). Finally, to confirm the robustness of this correlation, a secondary analysis that involved fitting a sigmoid to the data was used as an alternative method for extracting the pace parameter in the FORCED task.

Two important supplementary modeling analyses were conducted to test alternative interpretations of the results of the FORCED task: To further confirm the presence of intermediate movements, a mixture model analysis was conducted, where we fit combined uniform and Von Mises distributions to pooled distributions of reaches (Figure 4.12; Fan and Turk-Browne, 2016). Second, to test an alternative interpretation of our results (“response substitution,” see below; Cisek and Scott, 1999) a neural model was simulated which modeled movement speed and direction using cosine-tuned units (Figure 4.13). These analyses are discussed in further detail below.

4.4.2 Behavioral results

In the FREE task, participants were generally accurate on execution trials, with reach angles on the execution trials tracking the magnitude of the rotation imposed on the previous learning trial (Figure 4.10A; $t(31) = 30.57, p < 0.0001$). As predicted, RT on execution trials was linearly scaled by the previously-observed rotation magnitude (Figure 4.10B; $t(31) = 6.59, p < 0.0001$), in agreement with a mental rotation (MR) process (Georgopoulos and Massey, 1987; Bhat and Sanes, 1998).

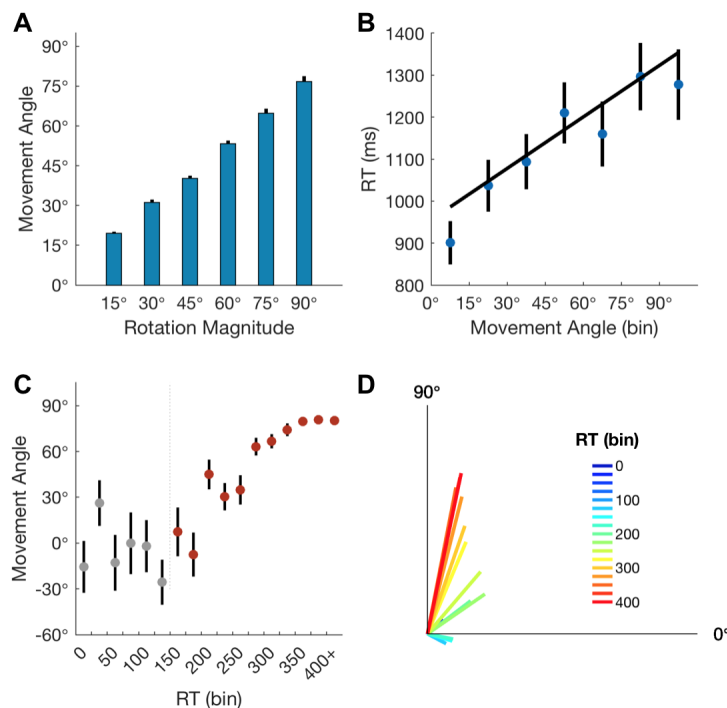


Figure 4.10: Behavioral results. (A) Absolute movement angles as a function of absolute rotation magnitude in the FREE task. (B) Mean of median RTs as a function of movement angle (binned by 15°). (C) Circular mean of movement angles as a function of RT (binned in 25 ms bins) in the FORCED task. Gray dots represent RT bins where movements were deemed as primarily random (see Methods). (D) Pooled vector representation of data in (C). Error bars = 1 s.e.m.

To track MR in the FORCED task, we first needed to identify the RT at which participants' movements "settled" on a direction that was clearly related to the presented target location. (At very short RTs, movements should be directed randomly since there is insufficient time to process the target.) To determine this "critical" RT, we looked at variance in reach direction in each successive 25ms RT bin. Circular variance significantly decreased ($t(31) = 2.28, p = 0.02$) from the 7th to the 8th RT bin (150-175ms), suggesting that at RTs over 150ms participants, began to make non-random movements. We note that this value is consistent with previous work (Haith et al., 2016). After this critical bin, mean reaching angles linearly increased with RT ($t(31) = 14.17, p < 0.0001$; Figure 4.10C). This result closely echoes the predictions of the rotation of a population vector in the motor cortex, and suggests an analog re-computing of the movement plan during the RT (Georgopoulos et

al., 1989). Moreover, the observed linear trend of intermediate movement directions (Figure 4.10C; Figure 4.10D shows data as circular stats) appear to demonstrate intermediate states of mental rotation; such states reflect a fundamental prediction of analog computations in general (Shepard and Metzler, 1971; Cooper, 1976).

Our conclusions assume that the FREE and FORCED tasks recruit the same mental rotation mechanism. To test this, we derived a mental rotation pace parameter (ms/deg) from each task. For the FREE task, we regressed each participants' execution trial data (Movement Angle X RT), and for the FORCED task we regressed each participant's full data set above the critical RT bin (RT X Movement Angle). The resulting regression slopes in the FORCED task were then inverted to match units. Pace parameters were strikingly similar between tasks ($t(31) = 0.19, p = 0.85$; Bayes Factor = 7.81 in favor of the null). Moreover, a significant correlation was found between participants' mental rotation paces on each task (Figure 4.11; $R(\text{pearson}) = 0.46; p = 0.008$; $R(\text{spearman}) = 0.41; p = 0.02$).

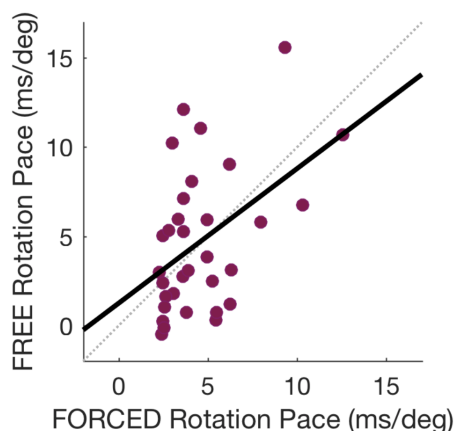


Figure 4.11: Correlation between tasks.

We note that RTs in the FREE task (Figure 4.10B) were well above those that produced equivalent movement directions in the FORCED task (Figure 4.10C). We believe this extra computation time is likely the result of a relatively low sense of urgency in the FREE task compared to the FORCED task. First, we note that the range of RTs observed in

the FREE task is comparable to the range observed in a similar study (Georgopoulos and Massey, 1987). Second, the regression analysis on the FREE task revealed no correlation across participants between the intercept of the regression, which reflects the putative non-rotation-related RT, and its slope, which reflects the mental rotation pace ($R(\text{pearson}) = 0.03; p = 0.87; R(\text{spearman}) = 0.09; p = 0.64$), suggesting that the “extra” RT in the FREE task is the product of processing that does not reflect MR.

4.4.3 Modeling results

There are two clear alternative explanations for the observed linear rise in mean movement angle in the FORCED task (Figure 4.10C, D): First, because participants made many random (i.e. uniformly distributed) movements at low RTs, and gradually made correct movements at higher RTs, a linear trend could appear as an artifact of averaging without any truly “directed” intermediate movements. Two control analyses contradict this explanation: First and most importantly, participants’ most frequent (mode) reach directions also displayed intermediate values, gradually increasing with RT (Figure 4.13 left, black dots; regression slope: $p < 0.01$). Second, the full distributions of reaching movements in the FORCED task show clear intermediate movements, and do not show any additional modes aside from near the 90° solution and near the -90° sign-flipped solution (Figure 4.12A).

For completeness, a mixture model accounting for random reaches and directed reaches was fit to the data. These modeling results also support the conclusion that directed reaches traveled through intermediate values.

The model was designed as follows: The probability density function of participants’ pooled reaching data at each RT bin was modeled as a mixture of two separate Von Mises distributions, one for correctly-signed reaches ($\mu_1 \geq 0^\circ$) and one for “sign flips” ($\mu_2 < 0^\circ$), as well as an additional uniform distribution capturing random reaches,

$$P = w_1 \frac{e^{\kappa \cos(x-\mu_1)}}{2\pi I_0(\kappa)} + w_2 \frac{e^{\kappa \cos(x-\mu_2)}}{2\pi I_0(\kappa)} + (1 - w_1 - w_2)U(-\pi, \pi); \quad (4.1)$$

where $I_0(\kappa)$ is the modified Bessel function of order 0, w is the weight given to each distribution ($w_1 + w_2 \leq 1$), κ is the concentration parameter (which is shared between the two pdfs), and $U(-\pi, \pi)$ is the uniform probability density function of the unit circle.

To further quantify the distributions of reach angle over the RT bins, we fit the data with two different versions of this mixture model. The first version, a “Free- μ ” model, allowed two mean parameters, one positive and one negative, to vary freely over RT bins (thus capturing mental rotation and sign flips). The second model, a “Fixed- μ ” model, fixed the mean parameters at $+90^\circ$ and -90° , which would be consistent with the *a priori* predictions of a discrete RC strategy (with swap errors).

Data were pooled across subjects in each RT bin and parameters were optimized separately in each bin using maximum likelihood estimation, minimizing the negative log likelihood with the MATLAB function *fmincon*. Fifty randomized starting parameter values were used for each fit to avoid local minima, and Aikake Information Criterion values (AICs) were summed over all RT bins after the critical 7th bin, or compared separately at each bin for model comparisons.

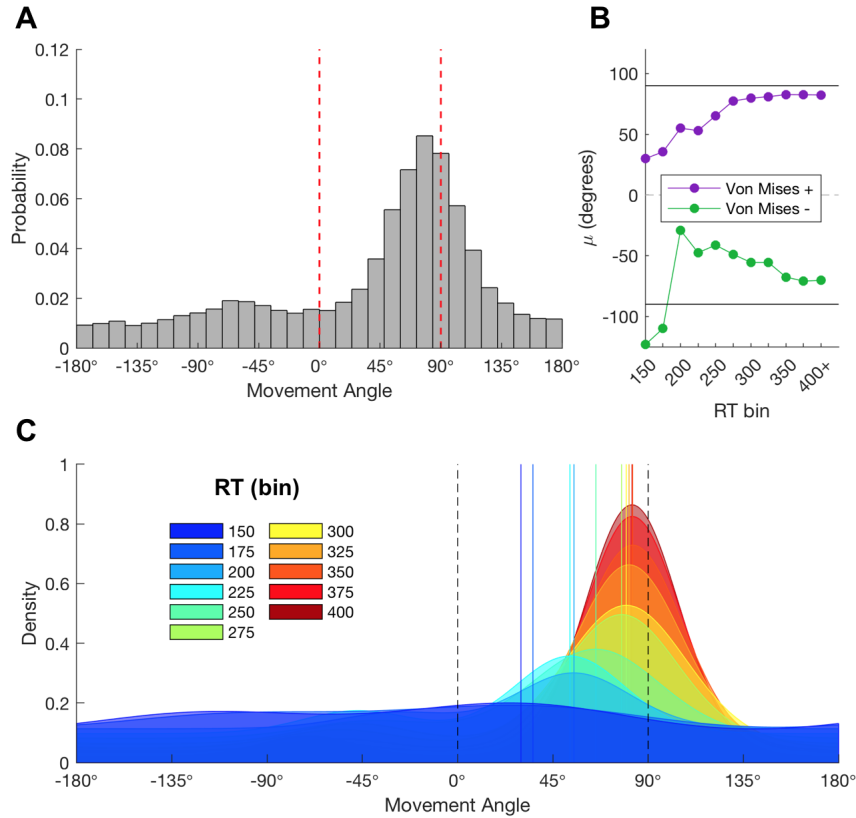


Figure 4.12: FORCED task full data distribution and model fits. (A) Full distribution of movement angles across all RTs in the FORCED task. (B) Fitted mean parameters from the “Free- μ ” mixture model (starting at the 150 ms bin) for directed reaches with a correct sign (Von Mises +, purple) and a flipped sign (Von Mises -, green). (C) PDFs of model fits in each RT bin. Vertical lines represent the fitted mean parameter (Von Mises +) for each corresponding RT bin.

Here we describe the results of the Free- μ model. As predicted, the mean parameters of the Free- μ model showed evidence for MR (Fig. 4.12B): Both the positive (purple) and negative (green) mean parameters gradually approached, respectively, +90° and -90° as RT increased. Linear regressions on the fit mean parameters revealed a significant positive linear trend for the positive μ parameter ($p < 0.001$), consistent with MR. Regression on the negative μ parameter, however, revealed no significant trend ($p = 0.38$), though we reasoned this result may be driven by the two deviant points (both > 1.5 sd from the mean) in the first two RT bins (Fig. 4.12B, green).

Figure 4.12C shows the full probability density functions of the best fit models in each RT

bin, with vertical lines representing the positive Von Mises mean in each bin. We emphasize that by accounting for random movements (with a uniform distribution), the mixture model offers further support for the mental rotation of planned, directed movements. Moreover, the apparent rotation of the negative Von Mises mean (Fig. 4.12B, green) suggests that sign flips themselves follow MR, albeit it in the erroneous direction. Critically, when the quality of fit of the Free- μ model is compared with the Fixed- μ model, we find that the Free- μ provides a far superior fit to the data ($\Delta\text{AIC} = 1189$); this large difference comports with MR, as the Free- μ model can more accurately capture intermediate movements.

Lastly, apparent mental rotation could appear as the result of “response substitution” between a prepotent response to the target location gradually being inhibited while the “rotated” response gradually replaces it (Cisek and Scott, 1999), essentially mirroring a kind of Stroop effect (Stroop, 1935). One prediction of response substitution would be the presence of bimodality in the data, where movements are made near the target (0°) in low RTs, and near the solution (90°) at high RTs, creating the illusion of mental rotation. However, we note that the distribution of reaching movements did not display such bimodality (Figure 4.12A), and distributions of the data across RT bins indeed show a shifting mean (Figure 4.12B, C).

Another model of response substitution could produce apparent mental rotation due to the averaging of simultaneously active motor plans (Cisek and Kalaska, 2005; but see Wong et al., 2017). Critically, this model of response substitution and our MR model make distinct predictions concerning movement *speed*: Response substitution predicts a clear slowdown of movement speed on trials where responses are averaged, whereas MR predicts a consistent movement speed during rotation. This difference can be captured in the length of the neural population vector, which has been linked to movement speed (Moran and Schwartz, 1999). We adapted a simple neural model introduced by Cisek and Scott (1999) to test this prediction. We also developed our own mental rotation variant of this model, and compared

the behavioral predictions of the two models.

The model of cortical directionally-tuned neurons is formalized as follows:

$$\frac{dx_i}{dt} = \kappa(-x_i + g_i E_i) \quad (4.2)$$

$$y_i = (x_i - \Gamma_i) + b_i \quad (4.3)$$

where y_i is the simulated activity of an M1 neuron i , and x_i reflects the input to that neuron. The parameter κ represents a time constant, g_i is a gain parameter, Γ_i is an activity threshold, and b_i is baseline activity (which is set to 0 in this model). During the simulations, if activity ever goes negative at any time point, it is reset to a value of 0. The rate of activity growth (Equation 4.2) is driven by the amount of excitatory input, E_i . In the simulations, cosine-tuned cells are given input at the start of the trial, with direction ϕ_j and magnitude a_j , where j indexes the details of the trial (i.e. target and goal directions),

$$E_i = \sum a_j [\cos(\theta_i - \phi_j) + s_i] \quad (4.4)$$

where θ_i is the cell's "preferred" direction and s_i is the offset of the tuning function. Preferred directions of the cells were uniformly distributed from 0° to 360° . All model parameters were set to the values reported in Cisek and Scott (1999). Moreover, the same number of neurons were simulated (300), though here the simulations were iterated 30 times (due to high noise in the parameter settings) to derive a range of responses.

Trials are simulated for each neuron, and the resulting population vector is computed as follows:

$$\vec{P} = \sum \vec{p}_i (y_i - b_i) \quad (4.5)$$

where $\vec{p}_i$ is the unit vector in each cell's preferred direction, which is summed across the population to yield $\vec{P}$. Much research has shown that movement speed is correlated

with the motor cortical population vector length (Moran and Schwartz, 1999); essentially, the more “confident” the population is of the planned movement direction, the faster the resulting movement tends to be. We converted the population vector lengths of both models to movement speeds (in cm/s) by scaling them by an arbitrary scaling parameter with a value of 0.76.

We modeled neural population activity as a result of either response substitution or mental rotation (MR) as follows: For response substitution, two distinct, independent inputs drive neural activity, one reflecting the target direction (e.g. 0°) and one reflecting the solution direction (e.g. 90°). For the first 150 ms, input represents solely the target direction, at a maximum magnitude of $a = 1$, and the magnitude of the solution direction (e.g. 90°) at the minimum $a = 0$. During the next 200 ms, these two magnitudes linearly switch values, with the target direction signal decaying to zero and the solution direction signal increasing to 1. The two signals then remain at constant values for the final 150 ms of the simulation.

For MR, only a single input is considered, and its direction is smoothly “rotated” during the RT. Like response substitution, for the first 150 ms the target direction drives the input at a magnitude of 1 (reflecting the subject processing the target location and anchoring where they need to rotate from). Then, the input angle (ϕ) linearly shifts from 0° to 90° for the following 200 ms, with no change in magnitude. The signal then remains at the solution direction for the remainder of time (150 ms). Lastly, for both models, an additional “cell recruitment delay” was drawn from a uniform distribution, $U(50 \text{ ms}, 150 \text{ ms})$.

These two input schedules reflect the logic of response substitution versus MR: In response substitution, a prepotent response to reach toward the target drives activity at first, and is slowly replaced by a new motor plan directed toward the solution. In MR, however, a single plan is directed toward the target and then smoothly rotated toward the solution.

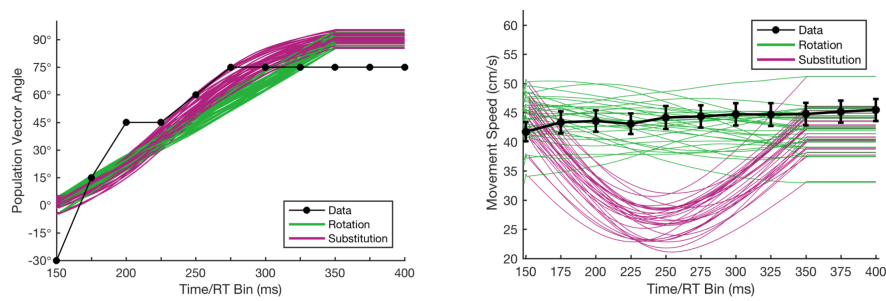


Figure 4.13: Neural model results. Left: Mode of reaching data from FORCED task (black dots), and simulated population vector angles from the MR neural model (green) and the response substitution neural model (purple). Right: Simulated population vector lengths (green, purple), and measured movement speed (black) over the RT bins. Error bars = 1 s.e.m.

As shown in Figure 4.13 (left), both models capture the broad shape of the angular data, though response substitution predicts a subtle sigmoidal shape relative to the strict linear shape of MR. Neither model appears to capture the logarithmic shape of the data (the mode reach directions are shown). Multiple green and purple lines show the results of 30 different simulations for each model. Critically, the two models make different predictions concerning movement speed (Figure 4.13, right): The MR model, because of its fixed input magnitude, predicts stable movements speeds over the course of mental rotation. However, the response substitution model predicts a clear “sink” in velocity during the substitution of responses, as the vector succumbs to “vote-splitting” between the target and solution (i.e. the population vector’s variance increases and its length decreases). Participants’ reach speeds (black dots) were relatively stable across mental rotation RT bins, as predicted by MR. The response substitution model provides a weaker picture of the movement speed data. Thus, this model provides additional evidence for the presence of MR in our FORCED task data.

Finally, to further test the presence of MR, we returned to the 2 X 2 experiment described in section 4.3. Using the average mental rotation pace from the FREE task, we can predict what the differences in RT between the 75° and 25° groups should look like when participants are using MR. As shown in Figure 4.14, the slope of RTs in the 12T groups at both the beginning and end of learning fall very close to the slope predicted by the FREE task, but

in the 2T group this is only evident in early trials, presumably when those participants are discovering the correct responses using a mental model.

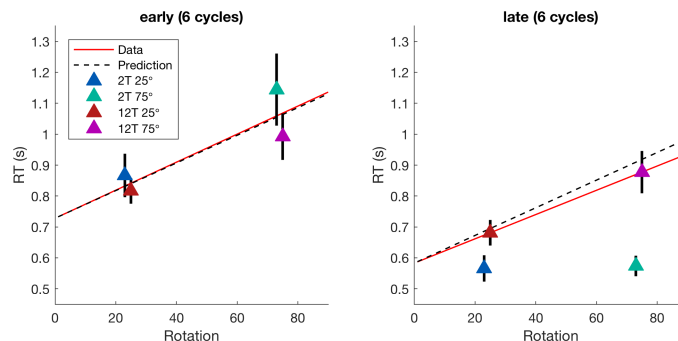


Figure 4.14: Mental rotation pace of FREE task predicts RTs in the standard, block-design VMR task. Left: Mean of median RTs in early learning (triangle markers) with a regression line, for visualization, on all four groups’ data (red), and the regression line predicted by the mental rotation pace gleaned from the FREE task using the same intercept (dashed black). Right: Same as left panel but with late learning data, with a regression lines for only the 12T groups (red).

4.4.4 Revealing response caching in a constrained-RT experiment

Having revealed the within-trial dynamics of MR (Figure 4.10), we now turn to response caching (RC). Based on the results of section 4.3, we predicted that altering the forced-RT procedure to only have two possible target locations would induce RC. Moreover, instead of seeing a gradual rotation of reach directions with increasing RT, we predicted that we would see a “step” function, consistent with the immediate recall of the response from working memory. Thus, instead of seeing intermediate movement directions, we predicted a strict bimodal distribution of responses at the two task solutions.

We used the identical Forced-RT task as before (Figure 4.9), but with only 2 possible target locations instead of 12, matching the set size manipulation of the 2 X 2 experiment ($n = 10$). We hypothesized that participants’ behavior in this context would reveal a bimodal distribution of reaches reflecting cached responses, but would not show evidence for mental rotation (i.e. via intermediate movements). (Note, we we did not include a FREE task for the 2-target condition, given that the FREE task itself requires a large variety of rotation

magnitudes and signs to test for a linear relationship between rotation and RT; under these conditions, a discrete RC strategy is not plausible given working memory constraints.)

With only 2 possible target locations, participants were able to fully counteract the rotation with very short RTs (Figure 4.15A). Examining the circular means of the movement angles reveals an abrupt jump from highly variable movements at short RTs ($< 225\text{ms}$) to consistent movements directed at the rotation solution ($+90^\circ$). Indeed, the pooled vector plot (Figure 4.15B) does not show the several highly-variable intermediate values seen in the circular means (Figure 4.15A), suggesting that those are likely the result of averaging of extreme values (e.g. -90° and $+90^\circ$; see below for model fitting). Furthermore, the full histogram of reaching directions (Figure 4.16A) is consistent with a bimodal response distribution: Participants' responses were concentrated near the solution ($+90^\circ$) and its opposite (-90°).

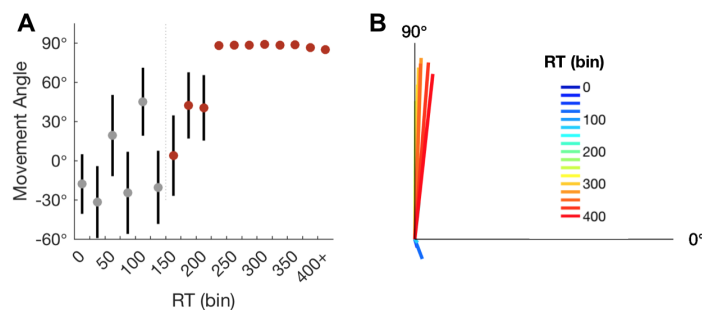


Figure 4.15: 2-target FORCED task results. (A) Circular mean of movement angles as a function of RT (binned in 25 ms bins) in the 2-target FORCED task. (B) Pooled vector representation of data in (A). Error bars = 1 s.e.m.

We note that given the “swap” errors observed in the 2T group in section 4.3, parsimony suggests that the negative responses in Figure 4.16A represent swap errors over “flip” errors, even though the location of each (-90°) is technically the same. Although we saw a subtle second mode at -90° in the 12-target FORCED RT experiment (Figure 4.12), that mode, given the number of targets, is only parsimoniously explained by flip errors (i.e. swap errors would not have a single mode, but would be uniformly distributed due to the large number

of responses).

To better quantify the distribution of reaches in the 2-target condition and compare the groups, we fit the Free- μ and Fixed- μ mixture models to the pooled data. Critically, the Von Mises mean parameters for the Free- μ model appeared to saturate immediately, not displaying the signature of mental rotation (Figure 4.16B, C). In fact, linear regression on the positive mean parameter revealed a slight negative slope ($p = 0.004$), which is in the wrong direction for mental rotation, and regression on the negative mean parameter revealed no significant trend ($p = 0.36$).

For our first comparison analysis, we analyzed the mode of pooled movement angles across the RT bins. Unlike the 12-target condition, we did not find a linear trend in the modes of the 2-target condition ($p = 0.20$). When the quality of fit of the Free- μ model is compared with the Fixed- μ model, we find that the Fixed- μ model provides a far superior fit to the data ($\Delta\text{AIC} = 1941$); this large difference comports with a discrete RC strategy, as there is no need for shifting mean free parameters in the absence of intermediate movements.

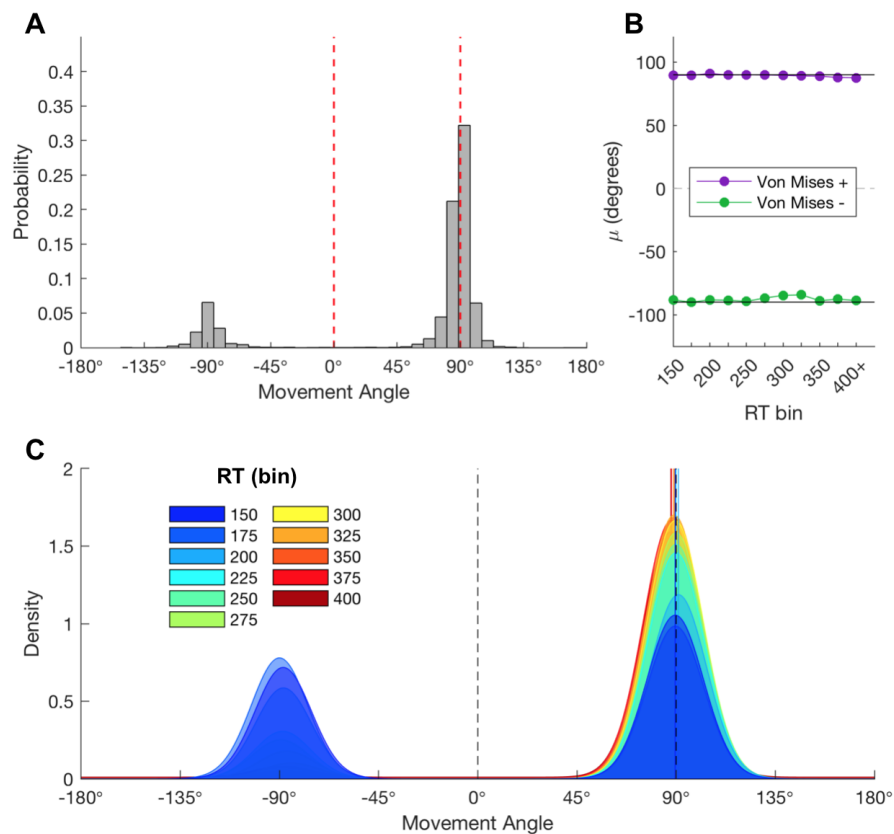


Figure 4.16: 2-target FORCED task full data distribution and model fits. (A) Full distribution of movement angles across all RTs in the 2-target variant of the FORCED task. (B) Fitted mean parameters from the “Free- μ ” mixture model (starting at the 150 ms bin) for directed reaches with a correct sign (Von Mises +, purple) and a flipped sign (Von Mises -, green). (C) PDFs of model fits in each RT bin.

Recall we found that the 12-target condition was better fit by the Free- μ model, consistent with using a parametric MR strategy, while the 2-target condition was better fit by the Fixed- μ model. However, for these analyses we summed AIC values across all model fits after and including the critical (7th) RT bin, perhaps inflating differences. Thus, we computed Δ AIC between models at each of these RT bins individually (Figure 4.17). The Free- μ model fit the 12-target group better at every RT bin (Δ AIC range: 6.92 - 276.13). In contrast, the Fixed- μ model fit the 2-target group better at every RT bin (Δ AIC range: 15.69 - 345.46). Thus, consistent with our predictions, participants in the 12-target condition were likely using a parametric MR strategy whereas participants in the 2-target condition were likely

using a discrete RC strategy.

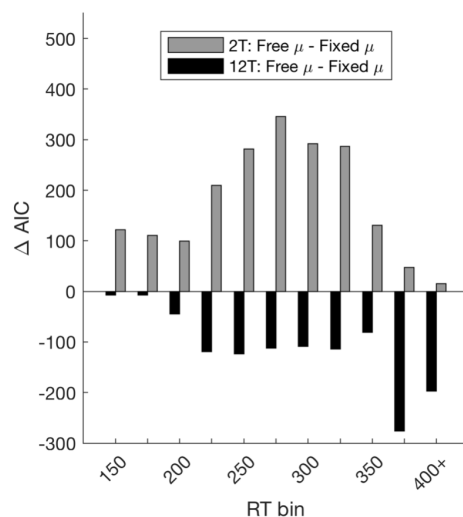


Figure 4.17: Model comparison. ΔAIC values over each RT bin between the Free- μ and Fixed- μ models for the 2-target (grey) and 12-target (black) FORCED RT experiments. Negative values constitute better fits for the Free- μ model, and positive values constitute better fits for the Fixed- μ model.

Taken together, these results show how two distinct working memory strategies (Georgopoulos and Pellizzer, 1995) can facilitate how an agent solves a motor learning problem (Figure 4.18) – in some cases, it is optimal to learn a rule, like MR, that can be applied generally. This strategy can be represented by an abstract transformation, like a rotation matrix. In other, simpler contexts, it may be optimal to only briefly use a rule to rapidly discover the optimal responses, and then store a look-up table of responses which can be quickly recalled and applied (RC). One future experimental direction could be to extend the length of training to see if this transition is inevitable, and investigate its time course.

The dissociation of parametric and discrete memory representations is a theme in several learning and memory domains. One example is human mathematical reasoning, where there is a strong behavioral and neural dissociation between numerical computations learned by rote and then recalled from semantic memory (“two times four = eight”), versus flexible parametric operations performed on a “mental number line,” represented in fronto-parietal

spatial working memory regions (Deheane and Cohen, 1997; Hubbard et al., 2005). Here we present an analogous dissociation, showing evidence for both the maintenance of discrete motor responses, and the parametric manipulation of such responses, in working memory.

I speculate that understanding the trade-off between these two strategies could be important for understanding how skills are eventually proceduralized: Does proceduralization require a progression from parametric to discrete representations? What are the selective advantages/disadvantages of one strategy or another? And what are the neural correlates of each?

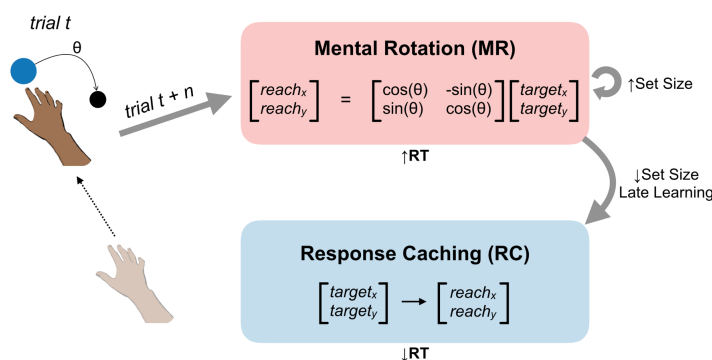


Figure 4.18: Schematic of learning strategies

4.5 Cerebellar and hippocampal roles in explicit motor learning

Understanding how an analog computation like mental rotation (MR) relates to motor behavior is difficult, as mental rotation is a covert mental activity. Sections 3.3 and 3.4 provided behavioral evidence supporting the theory that a rotation-like “analog” form of working memory aides in the learning of sensorimotor transformations. Another approach is to investigate if brain areas linked to motor behavior are implicated in such computations. Curiously, the cerebellum has been shown to be activated by mental rotation tasks (Stoodley, 2012), along with many other “motor-related” regions of the CNS (Zacks, 2008).

In this study we recruited individuals with spino-cerebellar ataxia (SCA), which causes wide-scale degeneration of cerebellar tissue. We hypothesized that SCA patients would show

disrupted mental rotation, while a non-analog form of working memory – working memory “scanning” – would be unimpaired. This hypothesis is based on both a previous empirical result and an intuition: First, cerebellar patients have been shown to have problems forming arbitrary stimulus response maps that aren’t directly cued (Spencer and Ivry, 2009); while mental rotation is not arbitrary *per se*, it is certainly not a direct S-R mapping. Second, and much more speculatively, the smooth rotating of a mental image may recruit a forward model, reflecting a link between dynamic mental imagery and sensorimotor prediction (Davidson and Wolpert, 2005). We hypothesize that neither of these points would apply to a discrete search in visual working memory.

Preliminary findings below support this prediction. Further, the results help explain a curious previous finding: Although SCA patients have impairments in both explicit and implicit VMR learning (Butcher et al., 2017), these impairments could be explained by the set size imposed during learning (i.e. the number of targets), further suggesting that MR is selectively impaired in SCA, but not RC.

In a second study, we investigate motor learning behavior in an individual, LSJ, who suffered almost complete bilateral loss of hippocampal and surrounding medial temporal lobe tissue due to viral encephalitis (Schapiro et al., 2014). We hypothesized that strategic adaptation to context changes in a visuomotor learning task would be difficult for this individual, as such experiences should rely on recalling previous episodes. This result would suggest that episodic knowledge has a role to play in motor learning as well. I also present preliminary evidence in support of this hypothesis.

4.5.1 The cerebellum and mental rotation

Here we tested the hypothesis that the cerebellum is recruited in parametric working memory tasks. Patients and age-matched controls performed both a classic alphabetic mental rotation task (Cooper and Shepard, 1973), as well as the Sternberg working memory task (“scanning”;

Sternberg, 1966). In the MR task (Figure 4.19, top), participants performed 216 trials. On each trial, an alphabetic letter (R, F, G, or J) was displayed on the screen in either a normal (e.g. R) or reflected (e.g. Я) orientation. Participants had up to 5 s to respond, using two designated buttons on a keyboard, choosing if the letter was normal or reflected. After their decision, feedback (correct vs. incorrect) was displayed for 1 s, then a fixation cross was shown for 2 s, and then the next trial began. Half of trials were normal letters and half were reflected. Critically, the displayed letters were rotated on each trial by 0° , 15° , 45° , 75° , 105° , or 135° , with the rotation sign randomized across trials.

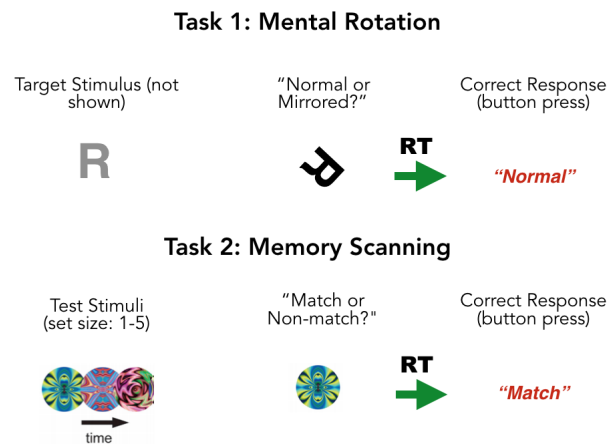


Figure 4.19: Task design for a standard mental rotation task (top) and a standard working memory scanning task (bottom).

In the memory scanning task (Figure 4.19, bottom), participants performed 100 trials. On each trial, they were shown a brief sequence of fractal images (1 s each; set size/load: 1-5 images), followed by a brief maintenance period (3 s), after which a new fractal image was presented that was either in the preceding sequence (“match”) or not (“non-match”). Participants had 5 s to decide. Feedback (correct vs. incorrect) was then displayed for 1 s, then a fixation cross was shown for 2 s, after which the next trial began. By varying the load (set size) across trials, we could make participants hold specific amounts of information in working memory on each trial. Original work using this task showed that RT is a linear function of load (Sternberg, 1966), much like mental rotation. However, this effect is

taken as evidence of a form of “working memory scanning” during the RT. Evidence from human behavior and monkey neurophysiology suggest that MR vs. scanning are distinct working memory computations – one requires the parametric manipulation of a representation, and the other a search through discrete representations (Georgopoulos and Pellizzer, 1995; Pellizzer et al., 1995).

Linear regressions were used to derive two parameters from participants’ data: An intercept and a slope (Figure 4.20A, B). The intercept reflects a baseline RT, or the amount of non-decision time (e.g. visual recognition, motor preparation, etc.). The slope parameter reflects the time of processing given the value of the independent variable, representing mental rotation speed for the rotation task, and mental scanning speed for the Sternberg task.

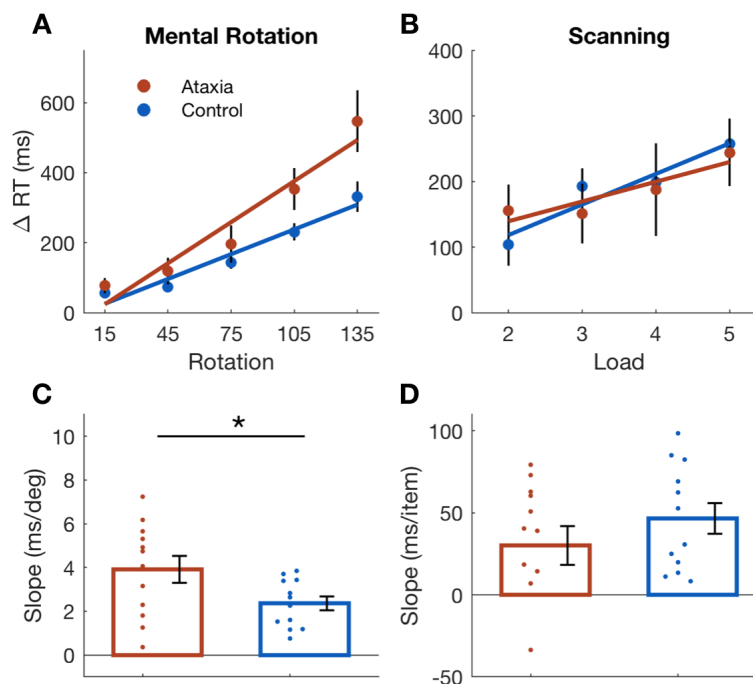


Figure 4.20: Slope results.(A) Ataxic patients (red) show a slower mental rotation slope than age matched-controls (blue), but similar scanning slopes (B). RT data are expressed as Δ RTs: Baseline RTs were subtracted off for each participant for the respective baseline conditions of each task (i.e. 0° rotation for MR, and load = 1 for scanning). (C) Bar graphs of individual slopes for MR, and scanning (D). * = $p < 0.05$. Error bars = 1 s.e.m.

SCA patients showed a higher slope parameter in the mental rotation task (Figure 4.20A,

C; $p = 0.035$), but not the Sternberg task (Figure 4.20B, D; $p = 0.28$), suggesting that parametrically manipulating a visual representation may draw on cerebellar processing. Importantly, we wanted to confirm that our effects were not driven by differences in raw response time; that is, we wanted to show that while slopes should not be correlated between tasks (i.e. MR and SB use different cognitive computations), baseline RT should be. First, SCA patients were slower than controls in non-decision time which was computed based on the baseline conditions of each task ($p < 0.001$). This is not surprising given the cerebellum's role in executing motor responses. Critically, however, baseline RTs showed a strong correlation between tasks ($r = 0.84, p < 0.001$), but not slopes ($r = 0.04, p = 0.82$), suggesting that non-decision time did not drive our main results (Figure 4.21).

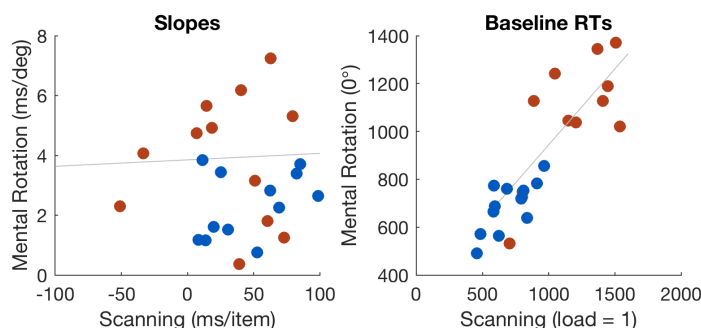


Figure 4.21: MR and scanning RT slopes (left) vs. RT intercepts (right)

Lastly, as described in Figure 4.22, although SCA patients were less accurate in both tasks (both $ps < 0.05$), these deficits appeared to be general effects that were not attributable to the independent variables of interest (i.e. rotation magnitude and load, respectively; Figure 4.22, center, right). However, further research will have to be done to explain these subtle but significant differences in accuracy.

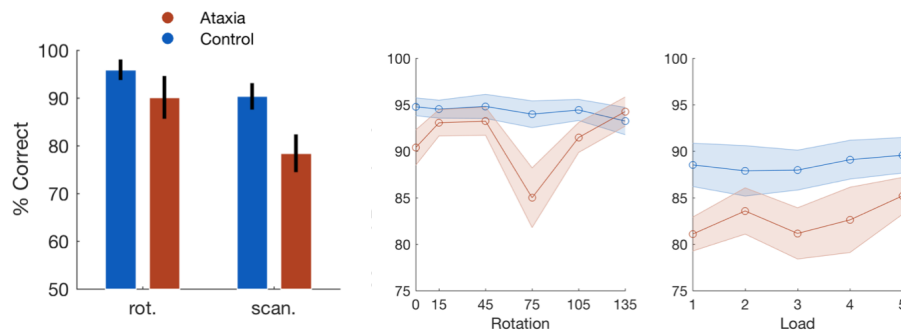


Figure 4.22: Accuracy results (left), and effects of independent variables on accuracy in MR (middle) and scanning (right). Shading and error bars = 1 s.e.m.

Upon closer inspection, these results could help explain a curiosity in recent work on cerebellar patients and visuomotor learning: In a recent study, Butcher and colleagues (2017) showed that SCA patients have deficits in both the explicit and implicit components of motor learning (see section 3 for a discussion of explicit and implicit processes in motor learning).

As shown in Figure 4.23, cerebellar patients both a) under-aimed in terms of their explicit motor learning (Figure 4.23B), and b) implicitly adapted less than age-matched controls (Figure 4.23C). The latter result replicates a classic finding from the motor learning literature (Martin et al., 1996), and is interpreted as resulting from damage to an internal forward model that drives adaptation (Taylor et al., 2010). However, the former result is a bit of a mystery – why would cerebellar damage affect explicit action selection?

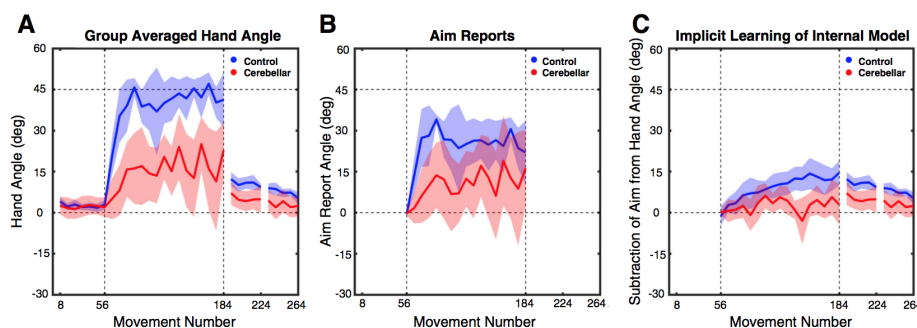


Figure 4.23: Cerebellar damage impairs both explicit and implicit learning (Adapted with permission from Butcher et al., 2017). (A) Hand angle learning curves for ataxic patients (red) and controls (blue). (B) Explicit learning and (C) implicit learning curves. Error bars = 95% confidence intervals.

One theory is that the cerebellum is important for arbitrary stimulus-response mapping. Spencer and Ivry (2009) showed that, on a sequence learning task, participants were impaired when attempting to learn indirect stimulus-response associations, but performed normally when the stimulus-response associations were direct, obvious mappings. It follows that the cerebellum may not be critical for habitual, trivial stimulus response behavior, like when a button is directly cued (Spencer and Ivry, 2009), or, in the case of reaching, when only a single response is correct and is repeated over and over.

Recall that in section 3.3, simple stimulus-response behavior (discrete response caching, or RC) was observed when the number of learning targets (set size) was low. However, when the set size was high, parametric mental rotation (MR), which requires the application of a novel abstract rule, was observed. In Butcher et al., 2017, the set size was eight targets, which would be taxing to stimulus-response working memory (Collins et al., 2012). Thus, it may be that MR was the predominant strategy used by the participants in Butcher et al., 2017. If so, observed SCA-linked deficits in explicit learning may be attributable to noisier, slower mental rotation by SCA patients described in Figure 4.20. If this hypothesis is correct, it follows that if SCA patients perform a visuomotor rotation experiment with a low set size (e.g. a single target), they may perform similarly to controls because they only need to cache a single response to successfully perform the task.

Indeed, in a preliminary data set from a new sample of SCA patients ($N = 10$), we observed no reliable impairment in explicit learning relative to controls when only a single target was learned (Figure 4.24; t-test on first cycle of explicit learning, $p = 0.08$). Overall learning was also comparable between controls and SCA patients, in both early learning ($p = 0.08$) and asymptotic learning ($p = 0.30$). Perhaps in this single-target context, SCA patients can rely on a simple response caching strategy. How to interpret the observed similarities in implicit learning, however, is not clear. I speculate that non-cerebellar-dependent implicit learning processes could explain why the patients show relatively normal implicit learning

in spite of cerebellar damage. One such process could be use-dependent learning, which is simply a learned bias that results from repeating a single movement direction (Verstynen and Sabes, 2011); another could be an operant bias that reinforces a rewarding movement (Huang et al., 2011).

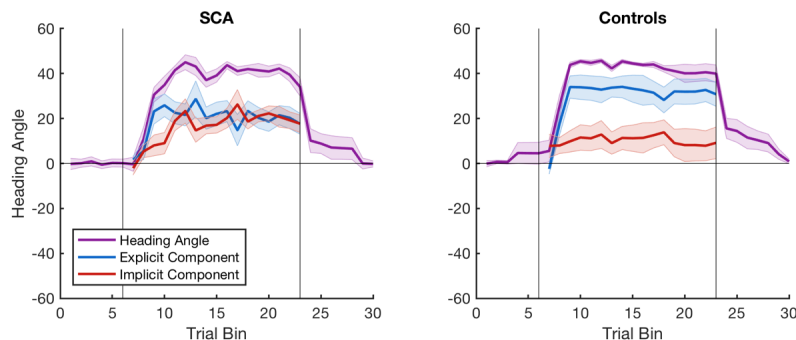


Figure 4.24: Results from 1-target experiment with SCA patients (left) and age-matched controls (right). Learning is depicted in terms of overall hand angle (purple), explicit learning (blue), and implicit learning (red). Error bars = 1 s.e.m.

One further speculation is that coordinating the parametric mental rotation of a visual percept may be analogous to coordinating a smooth movement, the latter being a known cerebellar function. In theory, cerebellar forward models could be more general-purpose than just being relevant for movement execution (Davidson and Wolpert, 2005). Future research could address this issue more directly, perhaps in the broader context of cerebellar roles in cognition (Stoodley, 2012).

4.5.2 The medial temporal lobe and flexibility in motor learning

One important prerequisite for intelligent action selection appears to be a knowledge of facts (Stanley and Krakauer, 2013). For example, adjusting your arrow to account for the wind involves recognizing a relevant context (the wind), and remembering that a particular context is best dealt with using a specific subset of actions (wind-adjusted shots). Here we tested this idea in VMR using a case study with a single extraordinary patient, LSJ.

In a series of seminal studies on the hippocampal patient HM, Milner and colleagues (Milner, 1962; Milner et al., 1968; Corkin, 1968; Shadmehr et al., 1998) revealed that HM retained the ability to learn novel motor tasks after his severe loss of medial temporal lobe (MTL) tissue. Specifically, HM showed clear signs of learning during mirror drawing (Milner, 1962), rotary pursuit (Corkin, 1968), and force field adaptation (Shadmehr et al., 1998). Given what we know about the neural substrates involved in motor learning, one hypothesis is that relatively spared circuits, such as the motor and premotor cortices, basal ganglia, and cerebellum, preserved HM's aptitude for motor learning. However, one interesting thread persisted through the various studies involving HM and his performance on motor tasks: His performance was always inferior to control participants, and he had to be explicitly instructed on how to do the task each day. If the hippocampus is not important for motor skill learning, why was this the case?

Here we present a brief case study with another hippocampal patient, LSJ. Due to an unfortunate case of viral encephalitis, LSJ sustained near-complete bilateral hippocampal loss, as well as damage to surrounding MTL cortex and the left anterior temporal lobe (Schapiro et al., 2014).

To test our hypothesis concerning the role of explicit context and motor learning, we asked LSJ to perform a “savings” experiment, where a rotational perturbation is learned, extinguished, and then relearned. Recent experiments have provided strong evidence that savings – a faster rate of re-learning than initial learning – in such tasks is likely the result of explicit memory processes (Haith et al., 2015; Morehead et al., 2015; Huberdeau et al., 2015). Moreover, savings has been attributed to rapid changes in explicit action selection in the face of context changes (e.g. rotation on versus rotation off; Morehead et al., 2015). We thus hypothesized that LSJ would show diminished savings relative to controls because she would forget which action selection policy to use. Our preliminary results support this hypothesis.

LSJ and 10 age- and handedness-matched controls (LSJ age: 66; mean control age: 68) participated in all three experiments. The control participants had normal or corrected-to-normal vision and hearing, no history of neurological disease, and were native English speakers. Control participants gave written, informed consent in accordance with the Princeton University IRB, and LSJ's consent was given by her legal caretaker. The Montreal Cognitive Assessment (MOCA) was used to screen control participants for cognitive deficits; the controls performed well on the MOCA with a mean score of 28 out of 30.

The task was performed on a robotic manipulandum (BKIN Technologies, Ontario Canada). Participants were seated in front of the robotic arm, which had a central position 15 cm from the participant's body. The robot sampled hand position, velocity, and force at 1 kHz. Stimuli were displayed on a horizontally-mounted mirror which reflected images from an LCD flat screen.

In general, visually-guided reaching tasks use small geometric shapes to represent cursors and targets (see previous chapters). However, in a pilot experiment, we found that such abstract stimuli were difficult for LSJ to interpret, and she often forgot the purpose of the experiment. Thus, we decided on a salient set of stimuli inspired by LSJ's past hobby of flying airplanes: In all tasks, the robotic arm controlled a small cartoon airplane that was supposed to contact the center of a cartoon runway on each trial (Figure 4.25). After this change, LSJ had no problem understanding the purpose of each task (i.e. "landing the plane").

"Savings" refers to the phenomenon where adaptation to a specific perturbation is faster on the second exposure, even after apparent extinction. In this task, the airplane was perturbed by a consistent -45° clockwise rotation during perturbation trials (Figure 4.25). Four targets were displayed at each of the four cardinal directions (0° , 90° , 180° , and 270°). Target position was pseudo-randomized in blocks of 4 trials. The task took place over a single session of 288 trials. The first 24 trials were baseline trials, where no perturbation

was applied. The next 80 trials were the first rotation block (R1), where the -45° clockwise rotation was applied on every trial. The following 80 trials were an “Extinction” block, where the rotation was removed. The next 80 trials were the second rotation block (R2), where the rotation was reapplied and savings was assessed. Finally, the last 24 trials consisted of a final washout block, where the rotation was again removed. A movement was considered complete when the hand passed an invisible ring around the center point with a radius of 10 cm. If a movement took longer than 2 s, a “too slow” message was displayed. The plane feedback was only presented at the end of the movement, signifying where the plane crossed the invisible ring. In a second experiment (Figure 4.27), the endpoint feedback was delayed by 1 s.

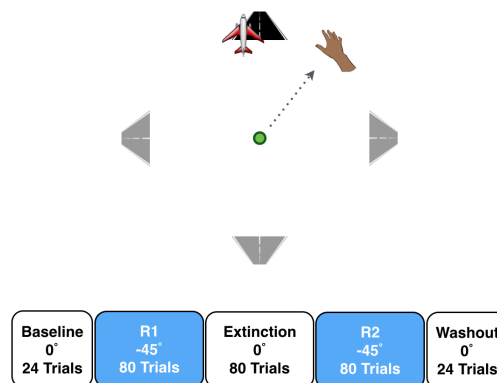


Figure 4.25: Task design. Participants tried to land a virtual plane on a virtual runway. Savings paradigm: The task starts with a baseline period with no rotation, followed by a first rotation block (R1), an extinction/washout period, a second rotation block (R2), and a final washout period.

Heading angle represented the participant’s reach angle relative to the target (e.g. a correct heading angle on a rotation trial would be $+45^\circ$), and was computed as the angle of the hand 2.5 cm into the full 10 cm reach. The savings metric was computed based on the method of Morehead et al., (2015): Savings was calculated by subtracting the mean heading angle for the first 12 trials in the R1 block from the first 12 trials of the R2 block. Importantly, those values were first baseline-corrected by subtracting the mean heading angle on the 10 trials preceding the corresponding block. Moreover, all reaches were baseline corrected based

on the initial baseline block. Finally, outliers were excluded from each perturbation block and the extinction block if the movement error exceed > 3 sd from the mean movement error in that block. Performance between LSJ and the age-matched controls was compared using z-scores.

As shown in Figure 4.26A, LSJ's initial learning (R1) phase was normal, if not faster than age-matched controls. This confirms that she can adapt to a visuomotor rotation, just as HM was shown to adapt to a force-field (Shadmehr et al., 1998). However, In the Extinction phase, LSJ did not fully wash out her learning, but the controls did. Critically, in term of the savings metric (learning rate in R2 relative to R1), LSJ had the lowest score relative to controls (Figure 4.26B, C; z -score = -1.65).

Savings in visuomotor learning tasks has recently been attributed to a “re-aiming” process, whereby the relearning context (R2) reminds the learner of the selection strategy they used in the initial learning phase (R1). This allows them to rapidly reduce errors in the second learning block without starting from scratch (Morehead et al., 2015). Consistent with that result, performance gains in the second rotation exposure load almost entirely on to the explicit component of learning and not the implicit component (Morehead et al., 2015). Moreover, savings is expressed by a learning component that requires high preparation time, but not a component that works at low preparation times (Haith et al., 2015), and savings can be observed even when the initial exposure to the first errors lasted for as little as 5 trials (Huberdeau et al., 2015). The latter result strongly suggests that savings is a function of a declarative memory of the errors themselves, as well as, perhaps, the method used to counter those errors.

I interpret the below results (Figure 4.26) as follows: At the onset of R2, LSJ may have experienced difficulty recalling her previous aiming strategy, and thus does not apply the correct behavioral policy to rapidly reduce her error. Moreover, the same process may also explain her “stickiness” in the Extinction block – while controls were likely rapidly re-aiming

in the opposite direction of their original aim when the rotation was turned off (Morehead et al., 2015), LSJ appeared to continue reaching in the manner she was reaching in the R1 block. Lastly, although we have no direct measure of the time course of implicit learning in this experiment, aftereffects between LSJ and controls were not significantly different (z -score = 0.25).

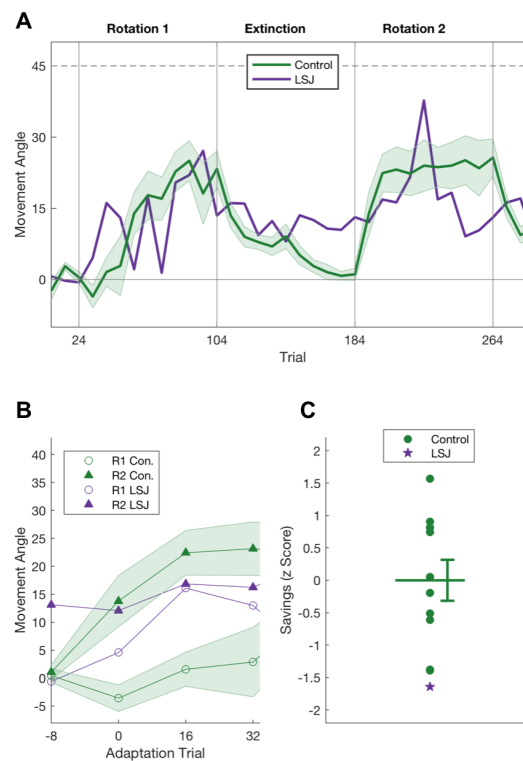


Figure 4.26: Disrupted savings in hippocampal patient LSJ. (A) movement angle across the task blocks in LSJ (purple) and age-matched controls (green). (B) Overlaid learning curves for the first rotation block (empty circles) and second rotation block (filled triangles). (C) Savings metric as z-score. Shading = 1 s.e.m.

One further question is whether LSJ is using explicit learning strategies at all. Given her aftereffects, it's reasonable to assume that she is implicitly adapting as normal. In a follow-up experiment, we delayed LSJ's feedback by 1 s to attenuate implicit learning (see section 3.3; Brudner et al., 2016). As shown in Figure 4.27, LSJ had extreme difficulty performing the task under delayed feedback, especially in the second rotation phase. She did, however, show some compensation for the rotation in R1 and R2 relative to baseline. Taken together, these

results suggest that a) most of her learning in the no-delay condition was likely driven by implicit processes, and b) while she may sporadically use working-memory-based strategies in the both conditions, contextual knowledge could be required to select and stabilize those strategies throughout the learning process.

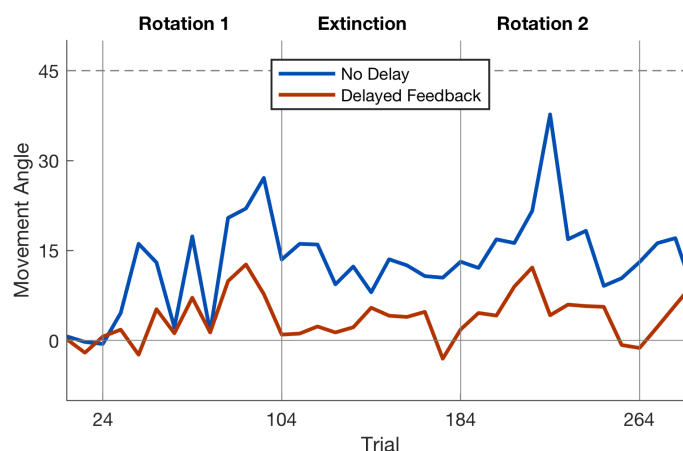


Figure 4.27: Effect of feedback delay on learning and savings in hippocampal patient LSJ.

To close this section, I present some brief anecdotal evidence from LSJ that complicates the story a bit, and speaks to the issue of learning timescales. The results of the above savings experiment (Figure 4.26) suggest that LSJ likely has trouble recalling the correct action selection policy in the face of a change of context. This effect is observed in a very short-term setting (i.e. an experiment that lasts $< \text{one hour}$). However, because of necessary piloting, LSJ actually used our robot on five separate occasions, often weeks apart.

On each visit after the first, we asked LSJ if she recognized the robot. Without fail, LSJ replied that she had never seen the machine and approached it with curiosity each time. However, and also without fail, following the initial training sessions she sat, leaned her head on the forehead rest, and grasped the robot arm without the need for instructions, in a manner that strongly implied familiarity with the machine. This anecdote is consistent with the idea that given enough exposure, a context-dependent, stimulus-response behavior can “sink in” in spite of MTL damage, perhaps reflecting a form of slow implicit learning

(Milner, 1962). Moreover, in this qualitative example, this kind of learning can not be easily attributed to simple error-based calibration (Shadmehr et al., 1998), but instead appears to involve the association of a complex context (e.g. the robot, experiment room, etc.) to a coherent set of actions (e.g. grasping the arm, resting the head on the cushion, etc.). How such complex associations can be learned without apparent encoding and recollection of explicit episodes is an open question, though certainly implicit reinforcement learning is a strong contender. However, I note that it would be misleading to call such behavior a “skill.”

4.6 General discussion

In this section I provided evidence supporting a role for controlled memory processes in motor learning. First, I showed that RT can be a useful measurement in the study of explicit motor learning strategies. Second, using both preliminary data and several novel experiments, I developed the hypothesis that these strategies can leverage either parametric or discrete working memory representations depending on the learning environment. Lastly, I provided evidence for how the cerebellum and medial temporal lobe may aid in explicit motor learning, with the former perhaps influencing complex parametric transformations in working memory, and the latter identifying contextual changes in the environment that can cue the recall of appropriate episodic memories.

How the hypothesized motor learning strategies play out in the long run, and whether they have any significant affects on the eventual proceduralization of a skill, are two vital open questions. One intuitive hypothesis is that any rule-based, parametric behavioral policy will eventually give way to purely “model-free” learning over time. Of course, the transfer of learning from goal-directed to habitual systems is a common theme in the human (Fitts and Posner, 1967) and animal literature (Costa, 2011).

Curiously, several studies on long-term visual mental rotation (using the task described in

Shepard and Metzler, 1971) find that although practicing mental rotation for weeks leads to an exponential reduction in overall RT, the mental rotation effect (i.e. RT as a linear function of rotation magnitude) decreases only subtly and does not go away (Leone et al., 1993; Wright et al., 2008). Critically, in the longest version of this task (Leone et al., 1993), a set size of 12 different objects and many rotation magnitudes was used, thus making a discrete RC (i.e., lookup table) strategy onerous. In a different study with long-term practice of visual mental rotation, only three stimuli and four rotation angles were used, and mental rotation effects significantly decreased after extensive training (Tarr and Pinker, 1989). Intriguingly, when novel orientations were introduced at the end of learning, the mental rotation effects returned, suggesting a dual-process model where subjects either retrieve an over-learned object from rote memory or perform a parametric computation (Tarr and Pinker, 1989; Provost et al., 2013). These results in the perceptual learning domain overlap considerably with our results in the motor learning domain, suggesting homologous mechanisms.

In terms of the look-up table strategy, the lines between explicit and implicit learning are blurrier. Indeed, a stimulus-response map is precisely how fully implicit, reflex-like habits are described (Costa, 2011). The key distinction to be made is between the active “scanning” of available responses in working memory versus the automatic production of those responses without said scanning. Indirect evidence for active scanning could come in the form of elevated RTs, swap errors (i.e. when the wrong response is called up from memory), or temporal decay (i.e. when the specificity of the response is weakened after brief delays). Normatively, these phenomena should not be common in the case of a habitual behavior. While these effects do appear in the experiments described above, further work is needed to delineate habitual from controlled stimulus-response learning. Does an explicit stimulus-response map evolve into an implicit one? Or are the two systems fully independent?

5 Conclusion

5.1 Proceduralization, knowledge, and cognitive contributions to skills

A key subject in the discussion of multiple learning processes in motor learning is the evolution of each process over time. Proceduralization, which describes the process of automizing a skill, is traditionally modeled as a gradual shift from declarative, explicit, learning strategies to purely implicit motor programs (Fitts and Posner, 1967; Anderson, 1982).

This idea of the “replacement” of controlled processes by implicit ones in long-term learning is the common view. In this view, learning is characterized by a gradual shift from a “cognitive” to “associative” to “automatic” stage (Fitts and Posner, 1967). This framework has recently been justified by experiments in cognitive neuroscience, which show how learning causes a shift in neural activity from anterior control regions (e.g. prefrontal cortex), to posterior associative cortices and subcortical regions (e.g. striatum; Dayan and Cohen, 2011).

Colloquially, evidence for this model of proceduralization is that experts are often unable to verbalize how they perform complex, precise motor behaviors. However, philosophical arguments on this issue make rather clear that verbalization should not be considered a prerequisite for knowledge (Wittgenstein, 1953; Stanley and Krakauer, 2013). Indeed, adjusting to a variety of novel contexts and explicit changes in environmental conditions is a common sign of an expert athlete (e.g. an expert archer adjusting to subtle wind gusts), even if they can’t necessarily verbalize every adjustment they make.

Perhaps a more accurate model of skill learning is one of a concerted interaction between knowledge, selection policies, and execution policies (Figure 5.1). Knowledge can be described as the set of propositional facts that tell the learner which subset of actions to select from (Stanley and Krakauer, 2013). In the savings experiment described in section

4.5, the re-introduction of the perturbation in the second learning phase lead to rapid re-learning in control participants, suggesting that they recalled which selection policy to use (see also Huberdeau et al., 2015; Morehead et al., 2015). In this sense, knowledge leads to improved selection of a selection policy – for example, “is my feedback being rotated, and should I thus mentally rotate my movement plan as I did before?” It follows that damage to the medial temporal lobe, as in patient LSJ, would severely diminish one’s capability to recall the relevant previous episodes and select the correct policy, which is indeed what we observed (Figure 4.25). Moreover, LSJ showed a catastrophic failure to learn when feedback was briefly delayed (Figure 4.26), and couldn’t even complete the task unless the stimuli were semantically related (i.e. a plane and a landing strip instead of different colored circles). (Even in the seminal case of HM, he had to be explicitly instructed about the mirror drawing task each time he performed it; see Stanley and Krakauer, 2013.)

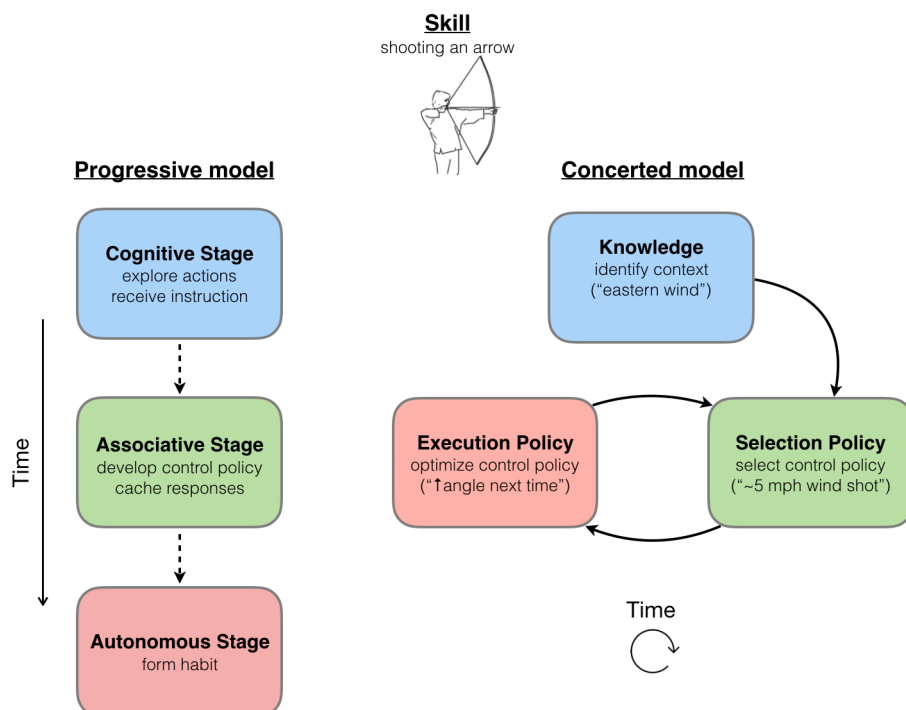


Figure 5.1: Models of skill learning

After the external context guides selection of the selection policy, the goal of the action

is specified and a high-value action is chosen. Selection can involve a discrete choice between responses (e.g. slot machine A versus slot machine B, or movement angle θ_1 versus movement angle θ_2) and/or a parametric computation in a continuous space of responses (e.g. mental rotation). Here, the selection policy can take the form of model-based planning, spatial reasoning, working-memory-based scanning of options, model-free habitual control, or some combination of these (Daw et al., 2005).

Finally, the execution policy specifies the transformation of the goal into the muscle commands that will achieve that goal. Changes in the execution policy are described by the implicit learning curve discussed in chapter 3. Changes in this process lead to one hallmark of implicit learning – aftereffects.

Critically, these three processes interact with each other in important ways. First, as mentioned above, knowledge interacts with the selection process by simplifying the selection problem; for instance, by inferring the underlying source of an action outcome (section 2) or by recognizing a familiar context (section 4.5). Selection interacts with execution by specifying which component motor programs to tinker with. This was modeled in section 3.2, where explicit action selection was shown to determine which movement representations are adapted during learning. In this framework, there is a cooperative feedback loop between these three processes.

Often it is the gradual improvement of the execution policy (recently termed “motor acuity”; Shmuelof et al., 2014) alone that people equate with skill learning. However, it is critical that parallel improvements in both the knowledge-based system and the selection system be incorporated into any theory of skill acquisition, as it is clear that true experts are adept in all three regards (Stanley and Krakauer, 2013). I believe the experiments outlined in this thesis show how even in simple tasks requiring discrete decision-making (section 2) or the learning of straightforward visuomotor mappings (sections 3 and 4), knowledge about the world plays an important role throughout learning.

5.2 Causal models and their role in learning

One idea that connects the various threads in this thesis is the idea of humans learning and applying a “causal model” of the world (Griffiths and Tenenbaum, 2009; Lake et al., 2017). Causal models appear to be a significant component of human learning (Griffiths and Tenenbaum, 2009). For instance, in a very successful recent model of written character identification, causality was operationalized by having the model treat the components of the characters as consequences of specific motor programs (i.e. pen strokes), reflecting a form of causal reasoning (Lake et al., 2015). This model performed as well as humans in rapid character classification, a notoriously difficult benchmark task.

To take a recent behavioral example of causal model building, Gershman and colleagues (2013) performed a simple fear conditioning experiment on rats: In one group, conditioning was followed by a standard extinction protocol, where no paired trials (i.e. tone + shock trials) were present. A second group experienced “gradual” extinction, where the probability of a paired trial was slowly reduced over time. In the former group, canonical spontaneous recovery was observed (Pavlov, 1927), where fear responses returned when the animal was reintroduced to the conditioned stimulus after a delay. However, spontaneous recovery was abolished in the gradual group, even though they experienced a greater total number of paired trials. This was taken as evidence for the updating of a causal model – by blurring the boundary between contexts (paired context versus unpaired context), the rats in the gradual group were made to think that a single world slowly changed, and that tones no longer predicted shocks. However, the standard extinction group may have formed a new model of the world when faced with the strong contextual shift between learning and extinction. In this case, the rat can be informally described as having two “worlds” in its head – one where tones cause shocks, and one where they do not. In this interpretation, spontaneous recovery reflects a return to the most salient world model.

How does causal model building relate to the results described in this thesis? In section

2, the causal model participants appear to use is relatively straightforward, informally translated as a simple rule – “only update my value representation of a stimulus on trials where I successfully caused it to reveal a reward outcome.” Here, sensory outcomes are classified by their cause: “Did I miss the target or hit the target?” Indeed, this “causal” classification of outcomes is exactly what the winning model is doing (the “gating” model). As shown in section 2.3, this causal model appears to be powerful enough to influence how dopamine-dependent reward prediction errors are computed in the striatum, providing one example of how high-level processes can influence putatively implicit learning processes (Leong, Radulescu et al., 2017; Collins et al., 2017).

In section 3 we began to tease out a potential causal model driving explicit learning in sensorimotor adaptation tasks. First, the heuristic apparently driving explicit learning can be informally translated, “if I see a large error, I should adjust my subsequent movement plan by a proportional amount in the opposite direction.” This rule is essentially what the modeled “fast” process of learning is doing (Figure 3.4). This rule does not, however, require a causal model, as it does not specify the generative process that led to the observed sensory outcome. Critically, as shown in section 3.2, explicit learning appears to generalize broadly in this task (see also Heuer and Hegele, 2011), suggesting a more flexible learning rule that infers the structure of the perturbation (Bond and Taylor, 2017). In section 4, we got closer to revealing that underlying rule. Consider the observed mental rotation of movement plans in the VMR task: Here, the causal model could be loosely translated as, “there is a consistent rotation of my visual feedback – with a sign ($\pm$) and magnitude (θ) – about the origin as a function of my movement direction.” Indeed, participants are usually able to verbally report this kind of strategy after the task is completed. In this example, sensory feedback is interpreted in a causal framework, and, further, the learner could be said to rely on a *conceptual* understanding of rotations.

Finally, in terms of the look-up table strategy, it appears that no causal model is needed.

Indeed, if the environment is simple enough, learning and applying a causal model may not be worth the cost. However, as shown in Figure 4.15, participants in the low set size VMR task appear to initially use parametric mental rotation to discover the task solution, and then transition to the faster look-up table strategy later on. Thus, even in this simple case, a causal model can still be important in the earliest phases of learning, prior to a shift to a more simple S-R policy later on. Future experiments could test if the causal model truly goes away, or if participants flexibly recall it based on the task context.

While the idea of causal models was not a focus of these experiments, it may prove a useful framework for explaining how subjects learn what to learn, and could help explain why people choose to perform rather abstract computations (like mental rotation) in the first place.

5.3 Parallels between models of reinforcement learning and motor learning

Below I will discuss several examples from reinforcement learning (RL) that have interesting parallels with the models of motor learning discussed in sections 3 and 4.

In one now classic investigation of human reinforcement learning theory, Daw and colleagues (2005) proposed a framework where learning is driven by both a) *model-free* strategies, which promote the repetition of rewarded actions, and b) prospective *model-based* strategies, which leverage a learned model of the structure of the world to predict the outcomes of actions or action sequences. This distinction can be elegantly teased out in a simple “2-step bandit task,” where participants first make choices that take them to one of two states with their own sets of choices, and then make a second choice that may or may not lead to reward. In this task, model-based reasoning takes advantages of transition probabilities between first- and second-level states to learn an optimal policy, whereas model-free learning merely rewards the specific actions that led to rewarded trials.

There is a tempting parallel to draw between model-based and model-free RL and, respectively, explicit and implicit motor learning. One point of caution, however, is that implicit motor learning itself is partially attributable to cerebellar-driven calibration (Tseng et al., 2007), and such calibration appears to rely on an internal “forward model” (Wolpert et al., 1995). Putting this form of implicit model-based calibration aside (as we claimed to accomplished experimentally using delayed sensory feedback, see section 4), we are left to speculate how precisely model-based and model-free RL could be manifested in motor learning.

Collins and colleagues (Collins et al., 2012; 2014; 2017) have provided an elegant framework for thinking about the concurrent operation of working memory- versus model-free RL-mediated learning. They designed a simple S-R task where participants have to learn to associate one of three buttons with different visual stimuli (e.g. blue square = button 2, red square = button 1, etc.). Critically, the authors titrated the number of S-R pairs, or the set size, in each learning block. A modeling analysis revealed that the best model of behavior included both a working memory module and a model-free RL module that operate in parallel during learning. Critically, the lower the set size, the more the working memory module drives learning and vice versa. Moreover, in an fMRI experiment they showed that reward prediction error signals in the striatum, the hallmark of RL, were blocked in low set sizes (Collins et al., 2017).

These results suggest that there is a trade-off between controlled processes and model-free RL depending on the learning environment, even in very simple tasks. This model has a rather strong parallel with the data presented in sections 3 and 4: I speculate that the working memory components highlighted in this thesis are homologous to those highlighted in Collins et al. (2012), and that the model-free RL component described in Collins et al. (2012) is akin to the incremental reinforcement of actions that were happened upon by the specific working memory strategy used in the task, be it parametric mental rotation or discrete response scanning.

Huang et al. (2011) recently provided evidence for the recruitment of a model-free operant process in visuomotor learning. Using a design that carefully isolated individual learning processes that were differently sensitive to hand movement, visual error, and reward, they revealed an operant learning process which reinforced successful movement directions in the context of a rotational perturbation. This putative operant process could reflect the model-free RL learning that occurs alongside working memory-dependent, model-based action selection highlighted in Collins et al. (2012). An additional source of evidence for this kind of learning in the experiments presented here is the reduction in overall RT, error, and variance over learning, which are standard correlates of proceduralization (Newell and Rosenbloom, 1980; Willingham, 1998; Shmuelof et al., 2014).

Response caching (section 4), while not itself “model-based,” can be divided into two forms: a controlled, working-memory dependent S-R map, and an implicit set of cached values. The former can be described by a maintained look-up table in working memory that explicitly represents the contingencies of the task (à la Collins et al.), while the latter can be described as a form of “habitual control” (Daw et al., 2005). Given this division, the response caching I described in section 4 likely requires the same neural and psychological mechanisms as model-based control. Indeed, disruption of the dorso-lateral prefrontal cortex (a key substrate in working memory) using theta-burst TMS leads to a down-weighting of model-based RL (Smittenaar et al., 2013). Habitual S-R learning, on the other hand, is tied to reward prediction errors in the associative striatum, and is considered to be synonymous with implicit RL (Daw et al., 2005; Collins et al., 2012; Costa, 2011). This model-free form of RL appears to persist throughout learning, and is thought to be the primary process operating during asymptotic learning (Collins et al., 2012). One open question is if model-based processes themselves become proceduralized (Economides et al., 2015), or are merely replaced by model-free ones.

Future research should attempt to further marry the fields of motor learning and rein-

forcement learning, as the theoretical overlap and potential for mutual insights is clear. In this spirit, motor skill learning should not be relegated to an isolated “procedural memory” branch in the taxonomy of memory processes (Squire, 1986).

References

- Albus, J. S. (1971). A theory of cerebellar function. *Mathematical Biosciences*, 10(1), 25–61.
- Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, 9(1), 357–381.
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89(4), 369.
- Anguera, J. A., Reuter-Lorenz, P. A., Willingham, D. T., & Seidler, R. D. (2010). Contributions of spatial working memory to visuomotor learning. *Journal of Cognitive Neuroscience*, 22(9), 1917–1930.
- Ashby, F. G., Alfonso-Reese, L. A., Turken, U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, 105(3), 442–481.
- Ashby, F. G. & Maddox, W. T. (2005). Human category learning. *Annu. Rev. Psychol.* 56, 149–178.
- Badre, D. & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, 45(13), 2883–2901.
- Barto, A. G. (1995). *Models of information processing in the basal ganglia: Adaptive critics and the basal ganglia*. MIT Press.
- Bays, P. M. (2016). Evaluating and excluding swap errors in analogue tests of working memory. *Scientific Reports*, 6, 19203.
- Bédard, P. & Sanes, J. N. (2014). Brain representations for acquiring and recalling visual-motor adaptations. *Neuroimage*, 101, 225–235.
- Bhat, R. B. & Sanes, J. N. (1998). Cognitive channels computing action distance and direction. *Journal of Neuroscience*, 18(18), 7566–7580.
- Bogacz, R. & Gurney, K. (2007). The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Computation*, 19(2), 442–477.

- Bond, K. M. & Taylor, J. A. (2015). Flexible explicit but rigid implicit learning in a visuomotor adaptation task. *Journal of Neurophysiology*, 113(10), 3836–3849.
- Bond, K. M. & Taylor, J. A. (2017). Structural learning in a visuomotor adaptation task is explicitly accessible. *eNeuro*, 4(4), ENEURO–0122.
- Bornstein, A. M., Khaw, M. W., Shohamy, D., & Daw, N. D. (2017). Reminders of past choices bias decisions for reward in humans. *Nature Communications*, 8, 15958.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, (10), 433–436.
- Brudner, S. N., Kethidi, N., Graeupner, D., Ivry, R. B., & Taylor, J. A. (2016). Delayed feedback during sensorimotor learning selectively disrupts adaptation but not strategy use. *Journal of Neurophysiology*, 115(3), 1499–1511.
- Butcher, P. A., Ivry, R. B., Kuo, S.-H., Rydz, D., Krakauer, J. W., & Taylor, J. A. (2017). The cerebellum does more than sensory prediction error-based learning in sensorimotor adaptation tasks. *Journal of Neurophysiology*, 118(3), 1622–1636.
- Castro, L. N. G., Monsen, C. B., & Smith, M. A. (2011). The binding of learning to action in motor adaptation. *PLOS Comput Biol*, 7(6), e1002052.
- Chen, C. H., Fremont, R., Arteaga-Bracho, E. E., & Khodakhah, K. (2014). Short latency cerebellar modulation of the basal ganglia. *Nature Neuroscience*, 17(12), 1767–1775.
- Christou, A. I., Miall, R. C., McNab, F., & Galea, J. M. (2016). Individual differences in explicit and implicit visuomotor learning and working memory capacity. *Scientific Reports*, 6, 36633.
- Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., & Shenoy, K. V. (2012). Neural population dynamics during reaching. *Nature*, 487(7405), 51.
- Cisek, P. & Kalaska, J. F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: Specification of multiple direction choices and final selection of action. *Neuron*, 45(5), 801–814.

- Cisek, P. & Scott, S. H. (1999). An alternative interpretation of population vector rotation in macaque motor cortex. *Neuroscience Letters*, 272(1), 1–4.
- Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the stroop effect. *Psychological Review*, 97(3), 332.
- Collins, A. G. E., Ciullo, B., Frank, M. J., & Badre, D. (2017). Working memory load strengthens reward prediction errors. *Journal of Neuroscience*, 37(16), 4332–4342.
- Collins, A. G., Brown, J. K., Gold, J. M., Waltz, J. A., & Frank, M. J. (2014). Working memory contributions to reinforcement learning impairments in schizophrenia. *Journal of Neuroscience*, 34(41), 13747–13756.
- Collins, A. G. & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? a behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience*, 35(7), 1024–1035.
- Coltz, J. D., Johnson, M., & Ebner, T. J. (1999). Cerebellar purkinje cell simple spike discharge encodes movement velocity in primates during visuomotor arm tracking. *Journal of Neuroscience*, 19(5), 1782–1803.
- Cooper, L. A. (1976). Demonstration of a mental analog of an external rotation. *Perception & Psychophysics*, 19(4), 296–302.
- Cooper, L. A. & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In *Visual information processing* (pp. 75–176). Elsevier.
- Costa, R. M. (2011). A selectionist account of de novo action learning. *Current Opinion in Neurobiology*, 21(4), 579–586.
- Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, 19(1), 51–57.

- Cunningham, H. A. (1989). Aiming error under transformed spatial mappings suggests a structure for visual-motor maps. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3), 493–506.
- Curtis, C. E. & D’Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, 7(9), 415–423.
- Dalton, J. C. (1861). On the cerebellum, as the centre of co-ordination of the voluntary movements. *The American Journal of the Medical Sciences*, 41(81), 83–88.
- Davidson, P. R. & Wolpert, D. M. (2005). Widespread access to predictive models in the motor system: A short review. *Journal of Neural Engineering*, 2(3), S313.
- Daw, N. D. & Doya, K. (2006). The computational neurobiology of learning and reward. *Current Opinion in Neurobiology*, 16(2), 199–204.
- Daw, N. D., Gershman, S. J., Ben Seymour, Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans’ choices and striatal prediction errors. *Neuron*, 69(6), 1204–1215.
- Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12), 1704.
- Daw, N. D., O’Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, 441(7095), 876–879.
- Day, K. A., Roemmich, R. T., Taylor, J. A., & Bastian, A. J. (2016). Visuomotor learning generalizes around the intended movement. *eNeuro*, 3(2).
- Dayan, E. & Cohen, L. G. (2011). Neuroplasticity subserving motor skill learning. *Neuron*, 72(3), 443–454.
- Dayan, P. & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience*, 8(4), 429–453.

- de Brouwer, A. J., Albaghdadi, M., Flanagan, R., & Gallivan, J. P. (2017). Gaze behaviour during sensorimotor adaptation parcellates the explicit and implicit contributions to learning. *bioRxiv*, 237651.
- Dehaene, S. & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex*, 33(2), 219–250.
- Diedrichsen, J., Hashambhoy, Y., Rane, T., & Shadmehr, R. (2005). Neural correlates of reach errors. *Journal of Neuroscience*, 25(43), 9919–9931.
- Donchin, O., Francis, J. T., & Shadmehr, R. (2003). Quantifying generalization from trial-by-trial behavior of adaptive systems that learn with basis functions: Theory and experiments in human motor control. *Journal of Neuroscience*, 23(27), 9032–9045.
- Donders, F. C. (1868). Over de snelheid van psychische processen. onderzoeken gedaan in het physiologisch laboratorium der utrechtse hoogeschool. *Tweede reeks*, 2, 92–120.
- Economides, M., Kurth-Nelson, Z., Lübbert, A., Guitart-Masip, M., & Dolan, R. J. (2015). Model-based reasoning in humans becomes automatic with training. *PLoS Computational Biology*, 11(9), e1004463.
- Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, 113(28), 7900–7905.
- Fan, J. E. & Turk-Browne, N. B. (2016). Incidental biasing of attention from visual long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(6), 970.
- Fernandez-Ruiz, J., Wong, W., Armstrong, I. T., & Flanagan, J. R. (2011). Relation between reaction time and reach errors during visuomotor adaptation. *Behavioural Brain Research*, 219(1), 8–14.
- Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6), 381–391.

- Fitts, P. M. & Posner, M. I. (1967). *Human performance*. Oxford: Brooks/Cole.
- Fodor, J. A. (1975). *The language of thought*. Boston, MA: Harvard University Press.
- Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage*, 6(3), 218–229.
- Georgopoulos, A. P. & Massey, J. T. (1987). Cognitive spatial-motor processes. *Experimental Brain Research*, 65(2), 361–370.
- Georgopoulos, A. P., Lurito, J. T., Petrides, M., Schwartz, A. B., & Massey, J. T. (1989). Mental rotation of the neuronal population vector. *Science*, 243(4888), 234–236.
- Georgopoulos, A. P. & Pellizzer, G. (1995). The mental and the neural: Psychological and neural studies of mental rotation and memory scanning. *Neuropsychologia*, 33(11), 1531–1547.
- Gershman, S. J. (2015). Do learning rates adapt to the distribution of rewards? *Psychonomic Bulletin & Review*, 22(5), 1320–1327.
- Gershman, S. J., Jones, C. E., Norman, K. A., Monfils, M.-H., & Niv, Y. (2013). Gradual extinction prevents the return of fear: Implications for the discovery of state. *Frontiers in Behavioral Neuroscience*, 7, 164.
- Ghez, C., Favilla, M., Ghilardi, M. F., Gordon, J., Bermejo, R., & Pullman, S. (1997). Discrete and continuous planning of hand movements and isometric force trajectories. *Experimental Brain Research*, 115(2), 217–233.
- Gläscher, J., Daw, N., Dayan, P., & Doherty, J. P. O. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, 66(4), 585–595.
- Green, C. S., Benson, C., Kersten, D., & Schrater, P. (2010). Alterations in choice behavior by manipulations of world model. *Proceedings of the National Academy of Sciences*, 107(37), 16401–16406.

- Griffiths, T. L. & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, 116(4), 661.
- Haber, S. N. (2003). The primate basal ganglia: Parallel and integrative networks. *Journal of Chemical Neuroanatomy*, 26(4), 317–330.
- Haith, A. M., Huberdeau, D. M., & Krakauer, J. W. (2015). The influence of movement preparation time on the expression of visuomotor learning and savings. *Journal of Neuroscience*, 35(13), 5109–5117.
- Haith, A. M., Pakpoor, J., & Krakauer, J. W. (2016). Independence of movement preparation and movement initiation. *Journal of Neuroscience*, 36(10), 3007–3015.
- Heuer, H. & Hegele, M. (2011). Generalization of implicit and explicit adjustments to visuomotor rotations across the workspace in younger and older adults. *Journal of Neurophysiology*, 106(4), 2078–2085.
- Hirashima, M. & Nozaki, D. (2012). Distinct motor plans form and retrieve distinct motor memories for physically identical movements. *Current Biology*, 22(5), 432–436.
- Huang, V. S., Haith, A., Mazzoni, P., & Krakauer, J. W. (2011). Rethinking motor learning and savings in adaptation paradigms: Model-free memory for successful actions combines with internal models. *Neuron*, 70(4), 787–801.
- Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, 6(6), 435.
- Huberdeau, D. M., Haith, A. M., & Krakauer, J. W. (2015). Formation of a long-term memory for visuomotor adaptation following only a few trials of practice. *Journal of Neurophysiology*, 114(2), 969–977.
- Huberdeau, D. M., Krakauer, J. W., & Haith, A. M. (2015). Dual-process decomposition in human sensorimotor adaptation. *Current Opinion in neurobiology*, 33, 71–77.
- Hyun, J.-S. & Luck, S. J. (2007). Visual working memory as the substrate for mental rotation. *Psychonomic Bulletin & Review*, 14(1), 154–158.

- Ito, M. (2006). Cerebellar circuitry as a neuronal machine. *Progress in Neurobiology*, 78(3), 272–303.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30(5), 513–541.
- James, W. (1890). *The principles of psychology*. New York: Holt.
- Kitazawa, S., Kohno, T., & Uka, T. (1995). Effects of delayed visual information on the rate and amount of prism adaptation in the human. *Journal of Neuroscience*, 15(11), 7644–7652.
- Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, 4(1), 47.
- Kosslyn, S. M., DiGirolamo, G. J., Thompson, W. L., & Alpert, N. M. (1998). Mental rotation of objects versus hands: Neural mechanisms revealed by positron emission tomography. *Psychophysiology*, 35(2), 151–161.
- Krakauer, J. W., Pine, Z. M., Ghilardi, M. F., & Ghez, C. (2000). Learning of visuomotor transformations for vectorial planning of reaching trajectories. *Journal of Neuroscience*, 20(23), 8916–8924.
- Krakauer, J. W. (2009). Motor learning and consolidation: The case of visuomotor rotation. *Advances in Experimental Medicine and Biology*, 629, 405–421.
- Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: Correcting a reductionist bias. *Neuron*, 93(3), 480–490.
- Krakauer, J. W., Ghilardi, M.-F., Mentis, M., Barnes, A., Veytsman, M., Eidelberg, D., & Ghez, C. (2004). Differential cortical and subcortical activations in learning rotations and gains for reaching: A PET study. *Journal of Neurophysiology*, 91(2), 924–933.

- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).
- Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338.
- Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*.
- Langdon, A. J., Sharpe, M. J., Schoenbaum, G., & Niv, Y. (2017). Model-based predictions for dopamine. *Current Opinion in Neurobiology*, *49*, 1–7.
- Leone, G., Taine, M. C., & Droulez, J. (1993). The influence of long-term practice on mental rotation of 3-d objects. *Cognitive Brain Research*, *1*(4), 241–255.
- Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017). Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron*, *93*(2), 451–463.
- Marcus, G. (2018). Deep learning: A critical appraisal. *arXiv:1801.00631*.
- Marr, D. (1969a). A theory of cerebellar cortex. *Journal of Physiology*, *202*, 437–470.
- Marr, D. (1969b). A theory of cerebellar cortex. *The Journal of Physiology*, *202*(2), 437–470.
- Martin, T. A., Keating, J. G., Goodkin, H. P., Bastian, A. J., & Thach, W. T. (1996). Throwing while looking through prisms. *Brain*, *119*(4), 1183–1198.
- Mattar, A. A. G. & Ostry, D. J. (2007). Modifiability of generalization in dynamics learning. *Journal of Neurophysiology*, *98*(6), 3321–3329.
- Mazzoni, P. & Krakauer, J. W. (2006). An implicit plan overrides an explicit strategy during visuomotor adaptation. *Journal of Neuroscience*, *26*(14), 3642–3645.
- McClure, S. M., York, M. K., & Montague, P. R. (2004). The neural substrates of reward processing in humans: The modern role of FMRI. *The Neuroscientist*, *10*(3), 260–268.

- McDougle, S. D., Boggess, M. J., Crossley, M. J., Parvin, D., Ivry, R. B., & Taylor, J. A. (2016). Credit assignment in movement-dependent reinforcement learning. *Proceedings of the National Academy of Sciences*, *113*(24), 6797–6802.
- McDougle, S. D., Bond, K. M., & Taylor, J. A. (2015). Explicit and implicit processes constitute the fast and slow processes of sensorimotor learning. *Journal of Neuroscience*, *35*(26), 9568–9579.
- McDougle, S. D., Bond, K. M., & Taylor, J. A. (2017). Implications of plan-based generalization in sensorimotor adaptation. *Journal of Neurophysiology*, *118*(1), 383–393.
- McDougle, S. D., Ivry, R. B., & Taylor, J. A. (2016). Taking aim at the cognitive side of learning in sensorimotor adaptation tasks. *Trends in Cognitive Sciences*, *20*(7), 535–544.
- Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience*, *16*(16), 5154–5167.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81.
- Milner, B. (1962). Les troubles de la memoire accompagnant des lesions hippocampiques bilaterales. In P. Passouant (Ed.), *Physiologie de l'hippocampe* (pp. 257–272).
- Milner, B., Corkin, S., & Teuber, H.-L. (1968). Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of HM. *Neuropsychologia*, *6*(3), 215–234.
- Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive hebbian learning. *Journal of Neuroscience*, *16*(5), 1936–1947.
- Moran, D. W. & Schwartz, A. B. (1999). Motor cortical representation of speed and direction during reaching. *Journal of Neurophysiology*, *82*(5), 2676–2692.

- Morehead, J. R., Qasim, S. E., Crossley, M. J., & Ivry, R. (2015). Savings upon re-aiming in visuomotor adaptation. *Journal of Neuroscience*, *35*(42), 14386–14396.
- Moutsiana, C., Charpentier, C. J., Garrett, N., Cohen, M. X., & Sharot, T. (2015). Human frontal-subcortical circuit and asymmetric belief updating. *Journal of Neuroscience*, *35*(42), 14077–14085.
- Newell, A. & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In *Cognitive skills and their acquisition* (Vol. 1, pp. 1–55). Routledge.
- Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology*, *191*(3), 507–520.
- Niv, Y., Edlund, J. A., Dayan, P., & O’Doherty, J. P. (2012). Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *Journal of Neuroscience*, *32*(2), 551–562.
- O’Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, *38*(2), 329–337.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.
- Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. London: Oxford University Press.
- Paz, R., Boraud, T., Natan, C., Bergman, H., & Vaadia, E. (2003). Preparatory activity in motor cortex reflects learning of local visuomotor skills. *Nature Neuroscience*, *6*(8), 882–890.
- Pellizzer, G. & Georgopoulos, A. P. (1993). Common processing constraints for visuomotor and visual mental rotations. *Experimental Brain Research*, *93*(1), 165–172.
- Pellizzer, G., Sargent, P., & Georgopoulos, A. P. (1995). Motor cortical activity in a context-recall task. *Science*, *269*(5224), 702–705.

- Peterson, L. & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58(3), 193.
- Petrides, M. (1985). Deficits on conditional associative-learning tasks after frontal-and temporal-lobe lesions in man. *Neuropsychologia*, 23(5), 601–614.
- Poggio, T. & Bizzi, E. (2004). Generalization in vision and motor control. *Nature*, 431(7010), 768–774.
- Provost, A., Johnson, B., Karayanidis, F., Brown, S. D., & Heathcote, A. (2013). Two routes to expertise in mental rotation. *Cognitive Science*, 37(7), 1321–1342.
- Pruim, R. H. R., Mennes, M., Buitelaar, J. K., & Beckmann, C. F. (2015). Evaluation of ICA-AROMA and alternative strategies for motion artifact removal in resting state fMRI. *Neuroimage*, 112, 278–287.
- Putnam, H. (1975). *Mind, language and reality (vol. 2)*. Cambridge: Cambridge University Press.
- Rand, M. K. & Rentsch, S. (2014). Gaze locations affect explicit process but not implicit process during visuomotor adaptation. *Journal of Neurophysiology*, 113(1), 88–99.
- Redding, G. M., Rossetti, Y., & Wallace, B. (2005). Applications of prism adaptation: A tutorial in theory and method. *Neuroscience and Biobehavioral Reviews*, 29(3), 431–444.
- Reppert, T. R., Lempert, K. M., Glimcher, P. W., & Shadmehr, R. (2015). Modulation of saccade vigor during value-based decision making. *Journal of Neuroscience*, 35(46), 15369–15378.
- Reynolds, J. N., Hyland, B. I., & Wickens, J. R. (2001). A cellular mechanism of reward-related learning. *Nature*, 413(6851), 67.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2), 225–237.

- Samejima, K., Ueda, Y., Doya, K., & Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science*, *310*(5752), 1337–1340.
- Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. *Journal of Cognitive Neuroscience*, *26*(8), 1736–1747.
- Schlerf, J., Ivry, R. B., & Diedrichsen, J. (2012). Encoding of sensory prediction errors in the human cerebellum. *Journal of Neuroscience*, *32*(14), 4913–4922.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.
- Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, *80*(1), 1–27.
- Schultz, W. (2013). Updating dopamine reward signals. *Current Opinion in Neurobiology*, *23*(2), 229–238.
- Schultz, W. (2015). Neuronal reward and decision signals: From theories to data. *Physiological reviews*, *95*(3), 853–951.
- Seidler, R. D., Kwak, Y., Fling, B. W., & Bernard, J. A. (2013). Neurocognitive mechanisms of error-based motor learning. *Advances in experimental medicine and biology*, *782*, 39–60.
- Seo, M., Lee, E., & Averbeck, B. B. (2012). Action selection and action value in frontal-striatal circuits. *Neuron*, *74*(5), 947–960.
- Shadmehr, R., Brandt, J., & Corkin, S. (1998). Time-dependent motor memory processes in amnesic subjects. *Journal of Neurophysiology*, *80*(3), 1590–1597.
- Shadmehr, R. & Mussa-Ivaldi, F. A. (1994). Adaptive representation of dynamics during learning of a motor task. *Journal of Neuroscience*, *14*(5), 3208–3224.
- Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annual Review of Neuroscience*, *33*(1), 89–108.

- Sharot, T., Kanai, R., Marston, D., Korn, C. W., Rees, G., & Dolan, R. J. (2012). Selectively altering belief formation in the human brain. *Proceedings of the National Academy of Sciences*, *109*(42), 17058–17062.
- Sharot, T., Korn, C. W., & Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. *Nature Neuroscience*, *14*(11), 1475.
- Sheahan, H. R., Franklin, D. W., & Wolpert, D. M. (2016). Motor planning, not execution, separates motor memories. *Neuron*, *92*(4), 773–779.
- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.
- Shepard, R. N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703.
- Shmuelof, L., Yang, J., Caffo, B., Mazzoni, P., & Krakauer, J. W. (2014). The neural correlates of learned motor acuity. *Journal of Neurophysiology*, *112*(4), 971–980.
- Smith, M. A., Ghazizadeh, A., & Shadmehr, R. (2006). Interacting adaptive processes with different timescales underlie short-term motor learning. *PLoS Biology*, *4*(6), e179.
- Spencer, R. M. & Ivry, R. B. (2009). Sequence learning is preserved in individuals with cerebellar degeneration when the movements are directly cued. *Journal of Cognitive Neuroscience*, *21*(7), 1302–1310.
- Squire, L. R. (1986). Mechanisms of memory. *Science*, *232*(4758), 1612–1619.
- Stanley, J. & Krakauer, J. W. (2013). Motor skill depends on knowledge of facts. *Frontiers in Human Neuroscience*, *7*, 503.
- Sternberg, S. (1966). High-speed scanning in human memory. *Science*, *153*(3736), 652–654.
- Stoodley, C. J. (2012). The cerebellum and cognition: Evidence from functional imaging studies. *The Cerebellum*, *11*(2), 352–365.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643.

- Sutton, R. S. & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.
- Tarr, M. J. & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, 21(2), 233–282.
- Taylor, J. A., Hieber, L. L., & Ivry, R. B. (2012). Feedback-dependent generalization. *Journal of Neurophysiology*, 109(1), 202–215.
- Taylor, J. A. & Ivry, R. B. (2011). Flexible cognitive strategies during motor learning. *PLoS Computational Biology*, 7(3), e1001096.
- Taylor, J. A., Klemfuss, N. M., & Ivry, R. B. (2010). An explicit strategy prevails when the cerebellum fails to compute movement errors. *The Cerebellum*, 9(4), 580–586.
- Taylor, J. A., Krakauer, J. W., & Ivry, R. B. (2014). Explicit and implicit contributions to learning in a sensorimotor adaptation task. *Journal of Neuroscience*, 34(8), 3023–3032.
- Taylor, J. A. & Thoroughman, K. A. (2007). Divided attention impairs human motor adaptation but not feedback control. *Journal of Neurophysiology*, 98(1), 317–326.
- Thach, W. T., Goodkin, H. P., & Keating, J. G. (1992). The cerebellum and the adaptive coordination of movement. *Annual Review of Neuroscience*, 15(1), 403–442.
- Thorndike, E. L. (1911). *Animal intelligence: Experimental studies*. New York: Macmillan.
- Thoroughman, K. A. & Shadmehr, R. (2000). Learning of action through adaptive combination of motor primitives. *Nature*, 407(6805), 742–747.
- Thoroughman, K. A. & Taylor, J. A. (2005). Rapid reshaping of human motor generalization. *Journal of Neuroscience*, 25(39), 8948–8953.
- Tseng, Y.-W., Diedrichsen, J., Krakauer, J. W., Shadmehr, R., & Bastian, A. J. (2007). Sensory prediction errors drive cerebellum-dependent adaptation of reaching. *Journal of Neurophysiology*, 98(1), 54–62.

- Verstynen, T. & Sabes, P. N. (2011). How each movement changes the next: An experimental and theoretical study of fast adaptive priors in reaching. *Journal of Neuroscience*, *31*(27), 10050–10059.
- von Helmholtz, H. (1924). *Treatise on physiological optics*. Rochester, NY: Optical Society of America.
- Vyas, S., Even-Chen, N., Stavisky, S. D., Ryu, S. I., Nuyujukian, P., & Shenoy, K. V. (2018). Neural population dynamics underlying motor learning transfer. *Neuron*, *97*(5), 1177–1186.
- Wei, K. & Koerding, K. (2010). Uncertainty of feedback and state estimation determines the speed of motor adaptation. *Frontiers in Computational Neuroscience*, *4*.
- Weiner, M. J., Hallett, M., & Funkenstein, H. H. (1983). Adaptation to lateral displacement of vision in patients with lesions of the central nervous system. *Neurology*, *33*(6), 766–772.
- Willingham, D. B. (1998). A neuropsychological theory of motor skill learning. *Psychological Review*, *105*(3), 558.
- Wimmer, G. E., Braun, E. K., Daw, N. D., & Shohamy, D. (2014). Episodic memory encoding interferes with reward learning and decreases striatal prediction errors. *Journal of Neuroscience*, *34*(45), 14901–14912.
- Wittgenstein, L. (1953). *Philosophische untersuchungen* (G. Anscombe, Trans.). Oxford: Oxford University Press.
- Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880–1882.
- Wong, A. L. & Haith, A. M. (2017). Motor planning flexibly optimizes performance under uncertainty about task goals. *Nature Communications*, *8*, 14624.
- Wright, R., Thompson, W. L., Ganis, G., Newcombe, N. S., & Kosslyn, S. M. (2008). Training generalized spatial skills. *Psychonomic Bulletin & Review*, *15*(4), 763–771.

- Wu, S.-W., Delgado, M. R., & Maloney, L. T. (2009). Economic decision-making compared with an equivalent motor task. *Proceedings of the National Academy of Sciences*, *106*(15), 6088–6093.
- Zacks, J. M. (2008). Neuroimaging studies of mental rotation: A meta-analysis and review. *Journal of Cognitive Neuroscience*, *20*(1), 1–19.