THE OPTIMALITY OF DECISION MAKING DURING MOTOR LEARNING

by

Joshua Brent Moskowitz

A thesis submitted to the Department of Psychology
in conformity with the requirements for
the degree of Master of Science

Queen’s University
Kingston, Ontario, Canada

(June, 2016)

Copyright ©Joshua B. Moskowitz, 2016
Abstract

In our daily lives, we often must predict how well we are going to perform in the future based on an evaluation of our current performance and an assessment of how much we will improve with practice. Such predictions can be used to decide whether to invest our time and energy in learning and, if we opt to invest, what rewards we may gain. This thesis investigated whether people are capable of tracking their own learning (i.e. current and future motor ability) and exploiting that information to make decisions related to task reward. In experiment one, participants performed a target aiming task under a visuomotor rotation such that they initially missed the target but gradually improved. After briefly practicing the task, they were asked to select rewards for hits and misses applied to subsequent performance in the task, where selecting a higher reward for hits came at a cost of receiving a lower reward for misses. We found that participants made decisions that were in the direction of optimal and therefore demonstrated knowledge of future task performance. In experiment two, participants learned a novel target aiming task in which they were rewarded for target hits. Every five trials, they could choose a target size which varied inversely with reward value. Although participants’ decisions deviated from optimal, a model suggested that they took into account both past performance, and predicted future performance, when making their decisions. Together, these experiments suggest that people are capable of tracking their own learning and using that information to make sensible decisions related to reward maximization.
Co-Authorship

All experiments were conducted under the supervision of Dr. Randy Flanagan. For the research presented in Chapter 2, I designed and carried out the optimality analysis, assisted with additional data analysis, took part in data collection, and provided input on design. This research was previously written up by independent project student Dan Gale (2015) towards completion of his undergraduate thesis. I took part in all aspects of the research presented in Chapter 3. Dr. Daniel Wolpert developed the optimality and fitting models used in that chapter.

Acknowledgements

I want to thank my thesis supervisor Dr. Randy Flanagan for his support and encouragement these past two years. From the beginning of this project we have attempted to explore new territory in our field of motor learning, and we now have a foothold on this exciting new area of research. I would also like to express my gratitude to Dr. Daniel Wolpert for his invaluable contributions to this project; especially to the amount of work he put in writing the MATLAB code that was used to analyze our data. Thank you to my committee members Dr. Stephen Scott and Dr. Gunnar Blohm for their advice and feedback. I would also like to acknowledge the people in my lab that have helped me see this thesis to completion. Included in this group are our post-doctorate fellows Dr. Jason Gallivan and Dr. Kevin Trewartha, who provided much needed expertise. I would especially like to thank undergraduate Dan Gale for laying the groundwork of this thesis, and undergraduate Mohammed Albaghdadi for assisting in data collection. Thank you to former graduate students Jonathan Diamond and Andreas Reichelt for showing me the ropes and to current students Ashley Bramwell & Tiger Ji for getting me to the finish line. Finally, I have a significant amount of praise for our programmer Martin York, who worked tirelessly to get my experiments up and running, and whose conversations are always enjoyable. Finally, I would like to thank my parents, Rosa and Willie Moskowitz for not just encouraging my studies, but fostering a love of reading and intellectual rigor that has stuck with me throughout my life.
Table of Contents

Abstract ................................................................................................................................. ii
Co-Authorship .................................................................................................................... iii
Acknowledgements ............................................................................................................ iv
List of Figures ..................................................................................................................... vii
List of Abbreviations .......................................................................................................... viii
Chapter 1: General Introduction ....................................................................................... 1
  1.1 Preamble ....................................................................................................................... 1
  1.2 Economic Decision Making in Humans ..................................................................... 2
  1.3 Decision making under risk – Movement planning ..................................................... 2
  1.4 Decision Making in Motor Tasks ............................................................................... 3
  1.5 Outline of Experiments ............................................................................................ 8
Chapter 2: Predicting future performance when learning a novel motor task ......................... 10
  2.1 Abstract ...................................................................................................................... 10
  2.2 Introduction ............................................................................................................... 11
  2.3 Materials and Methods ............................................................................................. 12
    2.3.1 Participants ........................................................................................................... 12
    2.3.2 Apparatus ........................................................................................................... 13
    2.3.3 Trial Procedure ................................................................................................... 14
    2.3.4 Experiment Phases ........................................................................................... 17
  2.4 Results ....................................................................................................................... 20
    2.4.1 Learning Analysis ............................................................................................... 20
    2.4.2 Decision Analysis ............................................................................................... 24
    2.4.3 Optimality Analysis ........................................................................................... 26
  2.5 Discussion .................................................................................................................. 32
Chapter 3: Reward optimization during motor adaptation ..................................................... 37
  3.1 Abstract ...................................................................................................................... 37
  3.2 Introduction ............................................................................................................... 38
  3.3 Methods ...................................................................................................................... 41
    3.3.1 Participants ........................................................................................................... 41
    3.3.2 Apparatus ........................................................................................................... 42
    3.3.3 Pilot Study ......................................................................................................... 43
    3.3.4 Stimuli .............................................................................................................. 44
List of Figures

Figure 1.1: Trommershäuser et al. (2006; 2009) design................................................................. 4
Figure 1.2: Trommershäuser et al. (2008) endpoint probabilities.................................................. 5
Figure 2.1: Apparatus used in the experiment. .............................................................................. 14
Figure 2.2: Top-down perspective of experiment workspace.......................................................... 15
Figure 2.3: Decision slider.............................................................................................................. 19
Figure 2.4: Learning data for three exemplar subjects................................................................. 21
Figure 2.5: Average reach error for four different 5-trial windows ............................................. 23
Figure 2.6: Average reward schedule selections across subjects for each learning horizon ....... 25
Figure 2.7: Average expected reward plots for each learning horizon in four exemplar subjects ... 27
Figure 2.8: Plot indicating the slope of the expected reward line and reward schedule selected for each learning horizon, for all subjects ........................................................................ 28
Figure 2.9: Average expected reward for each reward schedule and learning horizon ............... 30
Figure 2.10: Absolute differences in average expected reward between optimal and selected reward schedules as well as between Hit80Miss20 and selected reward schedules .......................................................... 31
Figure 3.1: Apparatus used for the experiment ............................................................................ 42
Figure 3.2: Example of a decision making trial ............................................................................ 45
Figure 3.3: Exemplar learning rates for four different subjects ................................................... 48
Figure 3.4: Target choices predicted by both an optimal and local optimal model compared to target choices made by the subject ................................................................................. 55
Figure 3.5: Target choices predicted by a lead, lag, and combined lead-lag model, compared to target choices made by the subject ................................................................................. 62
List of Abbreviations

ANOVA  Analysis of variance

HSD  Honest significance difference test

Hz  Hertz

kg  Kilo gram

m  Metre

mm  Millimetre

ms  Millisecond

M  Mean

N  Newton

N/m/s  Newton per metre per second

r  Correlation coefficient

s  Seconds

SD  Standard deviation

SE  Standard error

vBOT  Virtual reality robot
Chapter 1

General Introduction

1.1 Preamble

“Life is like playing a violin solo in public and learning the instrument as one goes on.”

- Samuel Butler

When humans engage in action they often do so in the attempt to accomplish some higher-level goal. In a sport such as golf, that goal would be to hit the ball into the hole using the least number of strokes. In everyday life, the goal may be to pick up a wine glass without knocking over the other drinks at the table. One major challenge in selecting a movement to accomplish these goals is that there are typically many different available movements that would suffice. In order to arrive at an optimal motor plan (i.e. one that best accomplishes our goal) we must first go through a decision making process (Wolpert & Flanagan, 2010; Wolpert & Landy, 2012).

In motor tasks, such decision making is complicated by the fact that the probability of successfully executing an action can vary as a function of a variety of factors (e.g., fatigue, environmental conditions, etc.) and can change dramatically with motor learning and adaptation. For example, the best choice of golf shot may vary with the wind conditions, and the best choice for a novice may be different than the best choice for a scratch golfer. Furthermore, the need to balance rewards and penalties creates an additional challenge for the decision maker. The choice can be between a high-risk, but potentially highly rewarding shot in golf (e.g., go for the green from a far distance), or a lower-risk, but less potentially rewarding shot (e.g., layup the ball before the green). The
prospect of arriving at an optimal decision seems faint given the number of factors one
has to incorporate into their decision making, and yet, professional athletes make these
types of decisions regularly and with apparent ease. This introduction will provide a
review of the human decision making literature, with an express emphasis on decision
making in the motor domain.

1.2 Economic Decision Making in Humans
Decision making in humans has been examined from both normative as well as
descriptive perspectives. Normative models of decision making such as expected utility
theory (Von Neumann & Morgenstern, 2007) are concerned with how people ought to
behave, looking at decisions in an economic framework that would attempt to maximize
financial reward and reduce variance. A descriptive model of decision making attempts to
understand how people actually behave. Prospect theory, proposed by Kahneman &
Tversky (1979), was one such attempt to provide a descriptive account of peoples’
decision making process. According to prospect theory, humans tend to make irrational
economic decisions that deviate from expected utility theory. People tend to overestimate
the occurrence of low-probability events, are risk adverse if there is a high probability of
gaining money, but are risk seeking if there is a high probability of losing money.

1.3 Decision making under risk – Movement planning
Selecting a motor plan is a case study of decision making under risk in the sense that we
select it without advance knowledge of the outcome (Kahneman & Tversky, 1984).
Although the motor system is capable of predicting the general consequences of
movements (Flanagan & Wing, 1997; Wolpert & Flanagan, 2001; Wolpert, Ghahramani,
& Jordan, 1995) there is an inherent uncertainty about the precise outcome of executing a
motor plan. This uncertainty stems from neural noise present in both the motor and sensory systems, but can also be due to the unpredictable external perturbations to the system such as wind when playing tennis (Wolpert & Landy, 2012). Another form of uncertainty arises when we attempt to learn a new motor skill. We may be unable to produce a desired movement outcome early in learning, but when we receive feedback from our motor performance, we are generally able to update our movement planning and control to reduce the amount of error on our next attempt (Krakauer & Shadmehr, 2006).

There is evidence that the motor system engages in decision making for even simple movement planning. For example, people factor in both biomechanical cost (Cos, Bélanger, & Cisek, 2011; Cos, Duqué, & Cisek, 2014) and task constraints (Cos, Medleg, & Cisek, 2012) when selecting and executing reaching movements and such factors are also accounted for when selecting corrective actions following an unexpected perturbation on the hand (Nashed, Crevecoeur, & Scott, 2012, 2014).

1.4 Decision Making in Motor Tasks
A major roadblock to determining whether humans are optimal in the realm of movement is the need to design a task that would involve maximizing gain in the face of competing motor choices. If individuals can select a movement that maximizes gain (associated with movement outcome) and do so consistently, then it could be argued that the motor system may be immune to the decision making biases of the cognitive system.

Trommershäuser and colleagues (2006; 2008) utilized a reaching task that required participants to consider costs and rewards in the face of uncertainty. The experiment was designed as follows. In an initial training phase, participants practiced executing rapid target-directed reaches to locations presented on a computer monitor placed in front of
them. Reaches had to occur rapidly (< 700 ms) and the target size was small, such that participants could not hit the target on every trial. This training phase provided participants information about their motor variability when reaching to the targets. In a subsequent phase of the experiment, participants were presented with visual stimuli that consisted of overlapping reward and penalty regions (Fig. 1.1).

![Diagram of target and reward regions](image)

Figure 1.1: **A**, an exemplar target used in Trommershäuser (2006; 2009). **B**, point values assigned to the reward, overlap, penalty, and non-reward regions.

Again, participants were required to reach the screen within 700 ms or face incurring a significant penalty. Participants received money when they successfully reached to the reward region, but lost money when they landed in the penalty regions or overlapping penalty and reward region.

Because participants could not guarantee hitting the reward region every time, due to motor variability, they were faced with the decision on where to aim such that they would maximize the amount of money they would receive. Selecting an optimal aim-point depended on participants being able to successfully incorporate information about their motor variability. By considering each participant’s variability, the authors could determine the optimal location they should aim towards in order to maximize reward and
compare this to the actual location the participant aimed towards (i.e., the average location across trials).

![Diagram showing movement endpoints and probability table]

Figure 1.2: A, movement endpoints for an exemplar participant taken from Trommershäuser et al., 2008. B, various probabilities of hitting a region of the target based on the exemplar subject’s distribution of movement endpoints.

The authors found that participants did not deviate significantly from optimal at the task and selected near-optimal aim-points from the very beginning of the reward phase, which is perhaps not overly surprising given that they practiced the task (Trommershäuser et al., 2006). Even more remarkable is that unlike the economic tasks that provide participants with explicit information about reward and probability, participants in this task were never given explicit information on the probability of successfully carrying out a potential movement plan; i.e., participants were not explicitly told their chances of hitting the reward or penalty regions. The results suggest that the motor system, which is engaged in movement planning, has access to this probability information.

Further studies have demonstrated near-optimal performance for both a similar speeded reaching task (Seydell, McCann, Trommershäuser, & Knill, 2008) as well as for a non-
speeded perceptual judgement task (Landy, Goutcher, Trommershäuser, & Mamassian, 2007). Moreover, when reach endpoint variability is artificially increased or decreased via altered visual feedback, participants adapt to the new variability and quickly adjust their aim-points so as to be optimal (Trommershäuser, Gepshtein, Maloney, Landy, & Banks, 2005). In another study, participants were asked to imagine moving from a start location to one of two targets that varied in size and distance, and to select the target that would result in the least movement time. Participants consistently selected the target that minimized movement time according to Fitts’ Law, suggesting that they had awareness of the speed-accuracy trade-off pertaining to reach aiming movements (Young, Pratt, & Chau, 2008).

Two different studies (Battaglia & Schrater, 2007; Faisal & Wolpert, 2009) constructed a task that examined trade-offs between viewing time and movement time. Participants had a fixed amount of time to view and reach to a target presented under conditions of visual noise. There was a direct trade-off between increasing viewing time, which allowed participants to obtain more accurate information about target location, and decreasing movement time, which increased signal-dependent motor variability. Participants in these studies scaled their viewing and movement times to the difficulty in viewing the target and moving towards it, such that they maximized their chances of contacting it. These results demonstrate that people can estimate visual and motor variability as a function of time, and use that information to select optimal viewing and movement durations.

However, there is evidence that in some circumstances the motor system performs sub-optimally at factoring in external reward. One study, using a task that was identical to a
study by Trommershäuser et al. described above (2006), recruited children between the ages of 6-11 years and found that they followed an overly risk-seeking strategy in their movements (Dekker & Nardini, 2015). Another experiment asked participants to select between a guaranteed reward and an option where they had to hit a small reward region surrounded by a penalty region (Nagengast, Braun, & Wolpert, 2011). Participants showed individual differences in risk-sensitivity, with some exhibiting risk-seeking behaviour (i.e., they valued high rewards even when the probability of obtaining the reward was low) and others exhibiting risk-neutral or risk-adverse behaviour. These results suggest that whereas some participants optimize mean reward, others place more value on variance of reward when making motor decisions. Using the same task as Trommershäuser (2006), Neyedli & Welsh (2014) found that when they shifted the penalty values of the target areas on a trial-by-trial basis, participants deviated from optimal performance. These authors also found that when they removed the initial training period present in the original study, participants were not initially optimal when reaching to the reward/penalty targets, and required a learning period before their movements were optimal (Neyedli & Welsh, 2013).

Finally, there have been some attempts made to construct motor tasks that would directly mirror the tasks used in studies on economic decision making. Wu et al. (2009) tested participants on both a classical economic decision between two lotteries as well as an equivalent decision between two motor lotteries (a rapid movement to a reward region surrounded by penalty regions). They found that in the motor task participants were more risk-seeking and tended to under weight small probabilities, while in the classical task participants over weighted small probabilities. Another study found that participants
showed similar performance levels in classical, motor, and arithmetic decision making tasks, but were faster to make decisions in the motor task (Jarvstad, Hahn, Rushton, & Warren, 2013).

The findings from these studies suggest that people, to some extent, are aware of the properties of their sensorimotor system, including visual and motor variability as well as the speed-accuracy trade-off, and can make use of this information to make decisions related to task performance. However, it seems that there are still clear individual differences in risk preference among variations in task.

1.5 Outline of Experiments
The goal of this thesis is to investigate the extent to which people are capable of making optimal decisions based on the performance of novel target-directed reaching tasks involving motor learning. In such tasks, optimal decisions—that maximize financial reward—requires that participants successfully predict their future performance based on their past and current performance. We performed two experiments to investigate this question. The first experiment examined whether participants could predict future task performance by requiring participants to select a reward structure tied to task performance partway through learning. The reward structure traded off points obtained simply for moving versus points for successfully hitting a target. Optimal performance depends on selecting a reward structure that best matched their future task performance. In the second experiment, participants could intermittently select target size, which varied inversely with the reward obtained for hitting the target, while adapting to a novel reaching task. Thus, selecting the target involved a trade-off between reward frequency
and reward value. Making optimal decisions in this task required the participant to predict future performance.
Chapter 2

Predicting future performance when learning a novel motor task

2.1 Abstract
It is unclear whether people can make optimal future decisions by incorporating knowledge of their own learning performance. We examined this question by engaging participants in a target-directed reaching task. After an initial practice phase, we applied a 45° visuomotor rotation to participants’ reaches. This exploration phase block was short (30 trials) such that participants only partially adapted to the rotation. In a subsequent decision phase, participants were told that they would next perform either 24, 150, or 216 more trials under the visuomotor rotation (harvest phase) and, for each trial block size, were asked to select a points ratio for target hits and misses, ranging linearly between 60:40 and 100:0. All participants were then informed that they had been assigned to a control group and performed 216 trials with an 80:20 ratio. We found that the selected point ratio increased with block size, suggesting that participants predicted their performance would gradually improve with practice. On average, we found that participants selected point ratios that were more optimal than the average (80:20) ratio they were assigned. We also found that participants made better choices for shorter block sizes and that participants whose task performance was at the extremes (i.e., very poor or very good) made more optimal selections. Taken together, these findings suggest that individuals can exploit information about their own learning in order to effectively guide decisions about future reward.
2.2 Introduction

A number of studies have demonstrated that people can take into account their own motor performance in order to optimize reward in the context of movement tasks. For example, when choosing where to aim when reaching toward a screen that displays reward and penalty regions, participants select an aiming direction that optimizes reward given their trial to trial movement variability (Trommershäuser et al., 2005, 2006; Trommershäuser, Maloney, & Landy, 2003a; Trommershäuser et al., 2008). However, little is known about how well people can take into account their future performance, based on past and current performance, when learning novel motor tasks. The ability to predict future performance in such situations is important because it would allow people to make an informed decision about whether to invest their time and energy in learning. Moreover, if an individual is required to learn a new task, accurately predicting future performance would enable them to make sensible decisions about assigning rewards (financial or otherwise) to success or failure in the task in the future.

To examine how well people can predict future performance on a novel motor task based on past performance, we constructed a simple motor learning task that required participants to adapt their target-directed reaches to a visuomotor rotation that altered the mapping between hand movement direction and the direction of a cursor controlled by the hand. During the initial exploration phase, participants performed 30 trials under the rotation. This brief exposure allowed participants to partially adapt their reaches to the rotation. Participants then took part in a decision phase in which they were told that in the subsequent harvest phase, they would be assigned to one of three experimental groups and would perform another 24, 150, or 216 trials under the rotation, depending on the
group. They were asked to select a points ratio, for hits and misses, ranging linearly between 60:40 and 100:0 for each of these three block sizes. Participants were aware that the more points they collected, the more money they would receive at the end of the experiment. Once they selected these ratios, they were informed that, in fact, for the harvest phase they had been assigned to a control group and performed a total of 216 trials with an 80:20 ratio. By examining their actual performance, we could assess how much reward they would have obtained for all possible points ratios.

Based on the hypothesis that participants were be able to predict their future performance based on their initial performance, we made the following three predictions. First, we predicted that that participants would select riskier (i.e. more points assigned to hits) points ratios as the number of harvest trials (i.e., length of the harvesting horizon) increased. Second, we predicted that participants would select points ratios that are matched with their own learning performance, with fast learners selecting riskier points ratios than poor ones. Third, we predicted that, for each of the three harvesting horizons, participants would select more lucrative points ratios than the ‘average’ ratio of 80:20. That is, we predicted that their choices would be in the optimal direction, relative to average ratio.

2.3 Materials and Methods

2.3.1 Participants

Eighteen right-handed participants from the Queen’s University student population were recruited for the study. Participants were between the ages of 19-27 years old ($M = 22.17$, $SD = 2.31$). One participant was excluded from data analysis due to them selecting a reward schedule that was 3 SD above the mean for all participants (100:0, 95:5, and
90:10 ratio of points for hitting and missing the target for the 24, 150, and 216 trial decisions respectively). Participants were compensated through a payment of $10, with the potential to earn more by receiving points during the experiment ($0.05 for every 100 points earned). Prior to the start of the experiment, it was ascertained that participants had no previous experience using the robotic device or performing a similar study. Participants provided informed consent before the experiment and were debriefed after the conclusion of the experiment. The experiment was approved by the Queen’s General Research Ethics Board and complied with the Declaration of Helsinki.

2.3.2 Apparatus

Participants grasped the handle of a two-dimensional planar manipulandum (vBOT; Howard et al., 2009) with their right hand to produce the required movements. The position of the handle was recorded at 1000 Hz. The visual stimuli used in the experiment, including a cursor controlled by the handle (filled orange circle, radius 5 mm), a start position (filled white circle, radius 5 mm) and a target (empty white circle, radius 10 mm), were projected by a computer monitor downwards onto a semi-silvered mirror that was visible to the participant and blocked vision of their hand (Fig. 2.1). To further prevent the participant from seeing their hand or arm, a curtain was fitted between the participant’s neck and the near side of the mirror. The height of the mirror was half way between the monitor and the center of the handle and therefore the visual stimuli appeared at the height of the center of the handle.
2.3.3 Trial Procedure

At the beginning of each trial, the vBOT applied forces to the handle so as to position it (and the cursor) over the start position. After a delay of 400 ms a target appeared at one of three pseudo randomly selected locations 100 mm from the start position and located at 0, +45, and -45 degrees relative to the start position (Fig. 2.2).
Figure 2.2: Top-down perspective of experiment workspace during a trial where the +45° target is selected (orange circle) and the visuomotor rotation is applied. Target position is determined pseudorandomly from among three possible (-45, 0, +45 degrees) positions. A force channel is applied 30 mm after movement onset to constrain out-and-back reaches to a straight line.

Participants were told to initiate a single continuous movement, out to the target and back towards the start position, as soon as the target appeared. Reaction time to the target presentation was strictly enforced in that participants were required to leave the start position no later than 350 ms after target presentation or else they received no points for that trial. This was done to minimize the possibility that, in trials in which a visuomotor rotation was applied (see below and Fig. 2.2), participants would engage a cognitive strategy of deliberating altering their trajectory to match the angle of visuomotor rotation (Fernandez-Ruiz, Wong, Armstrong, & Flanagan, 2011; Haith, Huberdeau, & Krakauer, 2015; Huberdeau, Krakauer, & Haith, 2015). The use of cognitive strategies is thought to result in more variable adaptation across trials (Fernandez-Ruiz et al., 2011) which may make the prediction of future performance based on initial adaptation more challenging.
An auditory cue sounded 500 ms after their handle left the start position and participants were instructed that the handle should arrive back at the start position in synchrony with this cue.

Participants were instructed to generate a straight out-and-back movement and not to make on-line corrections if the direction of the cursor did not match the direction of the target. However, to ensure that participants did not make corrections, once the cursor reached 30 mm away from the start position, a force channel was implemented in line with the vector joining the start position and the current position of the cursor. This channel constrained the handle to straight-line motion by applying force perpendicular to the direction of the channel if the handle deviated from the center of the channel.

Participants received substantial feedback on their motor performance during each trial. When the cursor reached 100 mm from the start position (i.e., the distance to the target), a ‘copy’ of the cursor was placed at the cursor position and remained on the screen for the rest of the trial. (If the cursor did not reach the 100 mm mark, the position of the cursor at the turn around point was extrapolated to the 100 mm mark and displayed.) This allowed participants to observe their angular reach error during and after the movement. Furthermore, at the end of the out-and-back movement text was displayed indicating if the movement was “Too Late” (reaction time > 350 ms), “Too Early” (reaction time < 50 ms), or—if the reaction time criteria were satisfied—whether the trial was a “Hit” (where any part of the cursor overlapped with any part of the target during the movement) or a “Miss”. Finally, at the start of each phase, a grid of empty white circles representing the number of trials to be performed (with 6 rows and N columns where N depended on the number of trials) was displaced and shown throughout the phase. One of these circles
changed colour at the end of each movement. The circles turned yellow if the reaction time to too short or too long, green if the trial was a hit, and red if the trial was a miss. This grid of circles enabled participants to monitor their performance throughout the phase.

2.3.4 Experiment Phases

Practice Phase.

Prior to the beginning of the experiment participants watched a short video tutorial that explained all the components of the experiment. The experimenter addressed any further questions. Following the video, participants first completed 30 practice trials without a visuomotor rotation and with an 80:20 ratio for hits and misses (i.e. participants received 80 points for each hit and 20 points for each miss respectively), or what we will refer to as a Hit_{80}Miss_{20} reward schedule. With knowledge of their performance on the first 30 practice trials (provided by the visual feedback, text feedback, and the grid of 30 circles), they then selected a reward schedule, ranging from for Hit_{60}Miss_{40} to Hit_{100}Miss_{0}, for an upcoming block of 102 practice trials using the decision slider (see below). We then implemented the chosen reward schedule for that block of trials.

Exploration Phase

In the exploration phase, participants performed 30 trials under a clockwise 45° visuomotor rotation. This brief exposure to the rotation allowed participants to partially adapt.
**Decision Phase**

After the exploration phase, participants selected a reward schedule for three separate harvesting horizons (i.e., three different length blocks of trials). Specifically, participants selected a schedule for a short (24 trials), medium (150 trials), and long (216 trials) horizon. Participants were informed that one of the three horizons would be implemented during the next harvesting phase of the experiment and that the reward schedule they selected for that horizon will be applied. After the reward schedules were selected for each learning horizon the experimenter informed the participant that, in fact, they had been assigned to a ‘control’ condition, in which they would perform the long horizon with a reward schedule of Hit\textsubscript{80}Miss\textsubscript{20}. This deception allowed us to obtain the same, large number (216) of trials for each participant and to control for any effect that choosing a high or low reward schedule might have on performance, while still getting unbiased decisions from our participants.

To select reward schedules, participants used a visual slider that they could control using the robot handle (see Fig. 2.3). At the top of the display, the grid of 30 coloured circles, which participants viewed at the end of the exploration phase, was re-presented so that participants could recall their performance. In addition, a set of 24, 150, or 216 unfilled circles, representing the number of trials to be performance in the subsequent harvest phase were displayed, providing participants with a visual sense of the number of trials.

Participants selected a reward schedule by moving the cursor over the slider and depressing a button on the top of the vBOT handle to ‘grasp’ the slider. By moving the handle to the left or right participants could select a reward schedule. Once the participant
was satisfied with the schedule they selected they confirmed their selection by releasing
the button and moving the cursor over to a text box with “OK” on it and depressing the
button again. The number of points for a hit and the number of points for a miss always
added up to 100. The slider is initially set in the middle to Hit$_{80}$Miss$_{20}$ (80 points for a hit,
20 points for a miss), and participants could adjust the slider by 1 point increments to as
low as Hit$_{60}$Miss$_{40}$, or as high as Hit$_{100}$Miss$_{0}$.

![Diagram](image)

Figure 2.3: During the decision phase, subjects select a reward schedule by clicking the
top button of the vBOT handle and moving it horizontally. The red line indicates points
assigned to missing the target, while the green line indicates points assigned to hitting the
target. Above the slider they are shown their performance on the 30 trials during the
exploration phase, and the number of upcoming trials they will perform.

**Harvest Phase**

In the final phase of the experiment, participants performed 216 trials under a clockwise
45° visuomotor rotation with a reward schedule of Hit$_{80}$Miss$_{20}$. Once they had completed
this phase, they were debriefed and paid.
2.4 Results

2.4.1 Learning Analysis

We first analyzed participants’ performance on the task by examining the angular error from their reaches. We measured the angle between the cursor position and the target at the point 100 mm away from the start location. In general, performance was near perfect during the practice phase of the experiment. Participants demonstrated close to +45 degrees of error during the initial trials of the exploration phase, but gradually adapted to the visuomotor rotation such that angular error was small by the end of the Harvest period.

Three exemplar participants (Fig. 2.4) showcase individual differences in performance during the exploration and harvest phases of the experiment. Participant P1 (Fig. 2.4A) was gradually adapted to the rotation and can only consistently hit the target (indicated by the green circles) towards the end of the harvest phase. Participant P5 (Fig. 2.4B) exhibited highly variable performance throughout learning with a number of fault trials (too slow or too early, see yellow circles), and finally participant P2 (Fig. 2.4C) was a fast learner who rapidly adapted to the rotation to produce a large number of hits throughout the harvest phase.

We fitted participant data using a single exponential function, consistent with previous studies that examined adaption to visuomotor rotations (Nikooyan & Ahmed, 2014; Zarahn, Weston, Liang, Mazzoni, & Krakauer, 2008).
Figure 2.4: A-C. Learning data for three exemplar participants. Trial outcome is colour-coded (green = hit, red = miss, yellow = too slow/ too fast). Vertical dashed lines indicate boundary of Exploration phase. Solid black line indicates fit of movement error to an exponential function of the form $ae^{bx} + c$, with equation provided. Fits were determined using MATLAB Curve Fitting Toolbox. A, participant who adapted slowly to the rotation and incurred a large number of misses. B, participant who was highly variable and incurred a large number of fault trials. C, participant who adapted quickly with few misses.

In order to determine how well participants adapted to the visuomotor rotation on average, we analyzed the mean error from the first five and last five trials of both the exploration and the harvest phase using a repeated measures ANOVA with a Greenhouse-
Geisser correction. There was a significant main effect of trial period, $F(1.69, 27.05) = 38.11, p < .001$, indicating that reach error changed between these blocks (Fig. 2.5).

Follow-up pairwise $t$ tests with a Bonferroni correction ($\alpha = .01$) revealed a significant reduction in mean error between the first ($M = 41.47, SE = 3.42$) and last ($M = 9.59, SE = 3.34$) five trials of the exploration phase, $t(16) = 6.17, p < .001$; the first ($M = 17.44, SE = 2.37$) and last ($M = 4.55, SE = .92$) five trials of the harvest phase, $t(16) = 5.48, p < .001$; the first five trials of the exploration phase and the last five trials of the harvest phase, $t(16) = 12.09, p < .001$; and the first five trials of the exploration phase and the first five trials of the harvest phase, $t(16) = 5.06, p < .001$. 
Figure 2.5: Average reach error for four different 5-trial windows. *** signifies that the comparison has $p < 0.001$. Error bars depict ±1 $SE$.

The difference in error between the last five trials of the exploration phase and the first five trials of the harvest phase was marginally significant, $t(16) = 2.78, p = .013$. This result was expected due to the gap in time between the completion of the exploration phase and the start of the exploration block during which participants’ memory of the adaptation would have decayed (Kitago, Ryan, Mazzoni, Krakauer, & Haith, 2013).

To assess how completely participants adapted to the visuomotor rotation, we compared the mean error across the last twenty trials ($M = 4.93, SE = .69$) of the Harvest phase to 0.
(zero error) using a one-sample t-test. A significant difference was found, \( t(16) = 7.14, p < .001 \), suggesting that participants did not completely zero out reach errors, a finding consistent with previous studies (e.g., Wigmore, Tong, & Flanagan, 2002). However, given that reach errors of \( \pm 8.53^\circ \) degrees would still result in hitting the target we can conclude that the mean error over the last twenty trials suggests that most participants were regularly hitting the target over this period despite not reducing error to zero.

2.4.2 Decision Analysis

We predicted that participants would select more risky (i.e., assigning more points to hits than to misses) reward schedules as the learning horizon size increased. We determined if this was the case by comparing reward schedule selections to horizon length using repeated measures ANOVA. Participants selected riskier reward schedules as learning horizon size increased, \( F(2, 32) = 15.76, p < .001 \) (Fig. 2.6A). Tukey’s HSD analysis found that participants selected lower points per hit for the short horizon (\( M = 63.24, SD = 4.71 \)) compared to the medium (\( M = 71.00, SD = 9.43 \)) and long (\( M = 75.24, SD = 13.47 \)) horizons, but that there was no difference between the two longer horizons.

Because reward schedule selections violated assumptions of normality and homogeneity of variance, Levene’s test, \( F(2, 48) = 6.81, p = .002 \), we conducted a non-parametric analysis using a Friedman test which showed that there were significant differences in reward schedule selection across learning horizons, \( \chi^2(2, N = 17) = 18.14, p < .001 \) (Fig. 2.6B). Pairwise Wilcoxon Signed Rank analyses with a Bonferroni correction (\( \alpha = .017 \)) indicated that there were significant differences between short and medium horizons (\( z = 3.09, p = .002 \)), short and long horizons (\( z = 3.08, p = .002 \)) and medium and long horizons (\( z = 2.50, p = .009 \)).
Figure 2.6: **A**, Average reward schedule selections across subjects for each learning horizon. Error bars depict ±1 SE. **B**, Individual reward schedule selection for each subject depicted by points, with lines connecting the decisions made by the same subject. ** indicates $p < 0.01$ for **A**, Repeated measures ANOVA, and **B**, Friedman test.

We also examined whether, for each of the three harvesting horizons, there was a correlation between performance and the selected reward schedule. For each horizon, performance was measured as the number of hits during the harvest phase and the reward schedule was expressed as the points allocated to hits (between 60 and 100). Pearson correlations did not indicate a significant correlation for the short horizon, $r(15) = .02$, $p = .94$, the medium horizon, $r(15) = .21$, $p = .42$, or the long horizon, $r(15) = .43$, $p = .09$. We also assessed, for each horizon, performance in terms of the average reach error during the harvest phase. Again no significant correlation between performance and
points allocated for hits was found for the short horizon, $r(15) = .11$, $p = .68$, the medium horizon, $r(15) = -.20$, $p = .44$, or the long horizon, $r(15) = -.35$, $p = .17$.

2.4.3 Optimality Analysis

We conducted an analysis to determine the extent to which participants selected reward schedules that were optimal (i.e. would receive the greatest amount of points possible) for their performance across each harvesting horizon. Specifically, we looked at how participants performed on the trials corresponding to each horizon (e.g. the first 24 trials of the harvest phase for the shortest horizon) and computed the points they would have received for every possible reward schedule from Hit$_{60}$Miss$_{40}$ to Hit$_{100}$Miss$_{0}$ (i.e. expected reward). By dividing the total number of points by the number of trials in the horizon, we determined the average expected reward for each horizon. Optimal reward schedules were bimodal, in that, for a given participant and horizon, either Hit$_{60}$Miss$_{40}$ or Hit$_{100}$Miss$_{0}$ resulted in the maximum average expected reward.

Figure 2.7 shows, for four exemplar participants, the expected reward as a function of points per hit for each horizon. A positive slope indicates that the participant performed sufficiently well over a particular horizon, and that selecting a Hit$_{100}$Miss$_{0}$ reward schedule would have resulted in them receiving the maximum amount of points possible. In contrast, a negative slope indicated poorer performance such that selecting a Hit$_{60}$Miss$_{40}$ reward schedule would have resulted in the participant receiving the maximum number of points possible for that horizon. Participant P2 (Fig. 2.7A) exhibited relatively good performance for all three learning horizons and selected riskier reward schedules as horizon length increased. Participant P1 (Fig. 2.7B) exhibited relatively poor performance across all three horizons but selected reward schedules that were optimal.
Participant P7 (Fig. 2.7C) showed mediocre performance for the medium and long learning horizons (slopes near zero), and poor performance for the short horizon. They selected riskier points schedules as learning horizon length increased. Participant P9 (Fig. 2.7D) showed poor performance across all three learning horizons, and yet, this subject chose riskier reward schedules as horizon length increased.

Figure 2.7: A-D. Each colored line indicates the amounts of points a participant would have earned on a single trial had a given reward schedule (points per hit) been applied to the trials included in that harvesting horizon (red = short horizon, green = medium horizon, blue = long horizon). Results are shown here for four exemplar participants. Colored dots indicate the corresponding reward schedules selected by the participant.

Figure 2.8 shows, for the three horizons of each participant, the relation between the slope of the expected reward line and the number of points assigned to a hit (note that a
black line connects each participant’s data points). The dashed horizontal and vertical lines indicate the average slope and point assignment for each horizon, respectively. Overall, slope increased with larger learning horizons, with negative slopes, on average, for the short (\( M = -14.31, SD = 12.35 \)) and medium (\( M = -2.24, SD = 12.23 \)) horizons and a positive slope for the long horizon (\( M = 1.33, SD = 11.36 \)). The Hit\(_{100}\)Miss\(_{0}\) reward schedule was only selected by one participant (P14), and only for the long horizon. This choice was optimal since the corresponding slope was positive. The Hit\(_{60}\)Miss\(_{40}\) reward schedule was often selected and this selection was, for the most part, optimal since the corresponding slope was negative.

Figure 2.8: Plot indicating the slope of the expected reward line and reward schedule selected for each learning horizon, for all subjects. Solid line connects points that belong to the same subject. Dashed horizontal and vertical lines indicate the average slope and
average reward schedule selection respectively, for each learning horizon. Grey bars indicate decisions that were optimal.

To determine whether participants, overall, selected reward schedules that were in the ‘optimal direction’, we compared, for each participant and horizon, the actual or selected expected reward, the optimal expected reward (resulting from selecting either Hit_{60}Miss_{40} or Hit_{100}Miss_{0} depending on the participant and horizon), and the expected reward had the participant selected the middle schedule of Hit_{80}Miss_{20} (Fig. 2.9). By examining the reward associated with the midpoint, we could assess whether participants tended to select a reward schedule in the direction of optimal (i.e., that earned them more reward than the reward associated with the middle reward schedule). As can be visually appreciated in Fig. 2.9, for all three harvesting horizons the selected expected reward was between Hit_{80}Miss_{20} and optimal.
Figure 2.9: Mean expected reward for the optimal, selected, and middle reward schedules for each harvesting horizon. Error bars represent ± 1 SE.

In order to determine whether participants’ reward schedule selections were in the direction of optimal we looked at the absolute difference in reward between the optimal and selected schedules, as well as the absolute difference in reward between the Hit\textsubscript{80}Miss\textsubscript{20} and selected schedules (Fig. 2.10). We evaluated whether these differences in reward were significant by carrying out one-sample t tests, using a test value of 0 with a Bonferroni correction (α = .008). The absolute difference in reward between the optimal and selected schedules were significantly different from zero for the medium, $t(16) = 3.65$, $p = .002$, and the long horizon, $t(16) = 3.15$, $p = .006$, but not for the short horizon,
$t(16) = 2.71, p = .02$. The absolute difference in reward between the Hit$_{80}$Miss$_{20}$ and selected schedules was significant for all three horizons (short, $t(16) = 5.51, p < .001$, medium, $t(16) = 4.08, p = .001$, and long, $t(16) = 3.51, p = .003$). These results indicate that, on average, participants selected reward schedules that were in the direction of optimal for the medium and long learning horizons, and reward schedules that were not reliably different than optimal for the short horizon.

![Figure 2.10: Absolute differences in average expected reward between Optimal and Selected reward schedules as well as between Hit$_{80}$Miss$_{20}$ and Selected reward schedules. Error bars represent ± 1 SE. ** represents $p < 0.01$, and *** $p < 0.001$.](image)
2.5 Discussion

The current study made use of a target-directed reaching task to investigate decision making during the course of motor learning. We wanted to determine if participants could predict their future task performance based on their performance to-date, and use that information to make optimal decisions about future reward. Specifically, we hypothesized that participants would be able to extract information from their initial learning parameters (e.g., rate, variability) and use that information to select a reward schedule that was optimal for a given harvesting horizon. Based on our hypothesis we made three predictions. First, we expected participants would select riskier (more points allocated to hitting the target) reward schedules as the harvesting horizon increased. Second, we expected that points ratio selection would correlate with task performance, with better performers choosing riskier ratios and poor performers choosing less risky ones. Third, we predicted that, for each harvesting horizon, participants would select a points ratio that was more rewarding than the average ratio of 80:20 (points per hit: points per miss). That is, we predicted that participants’ choices would be in the optimal direction, relative to average ratio.

The results of the study support our first prediction. Participants did indeed choose riskier reward schedules as the foraging horizon increased. This suggests that participants were capable of predicting that they would have more hits over the course of the longer horizons, which would offset points lost due to early misses. The fact that only one participant selected a Hit_{100}Miss_{0} reward schedule for the long horizon (Fig. 2.8, P14) suggests that participants were quite risk averse with their reward schedule selections. Despite a number of participants having positive slopes (more hits than misses) on the
medium and long horizons, they did not select the optimal reward schedule of \( \text{Hit}_{100}\text{Miss}_0 \). The reason for this could have been participants had difficulty calculating that future hits would fully offset the early loss in points from missing the target, or it could be an inability to fully predict the time course of learning.

In terms of the second and third predictions, the results are more unclear. While participants selected reward schedules that were in the direction of optimal (Fig. 2.10) their reward schedule selections did not significantly correlate with their performance on the harvest phase of the experiment. So while participants must have factored in their future performance to some extent while selecting reward schedules it does not seem to be the case that they were fully capable of extrapolating their future performance based on their performance in the exploration phase. Participants selected near-optimal reward schedules for the short, but not long or medium horizons, suggesting they had difficulty predicting performance across longer horizons.

One of the reasons for participants’ difficulty in predicting their learning might have been that subjects made use of a cognitive strategy when learning the task. Cognitive strategies are associated with highly variable performance (Taylor, Krakauer, & Ivry, 2014) and indeed this characteristic was present among a number of subjects (Fig. 2.4B). It is conceivable that this variability in performance makes it more difficult to extrapolate the underlying learning curve and thereby predict future performance. Furthermore, the presence of fault trials during the exploration phase suggests that despite our reaction time criteria, participants were (at least sometimes) performing reaches with high preparation time, allowing for use of a cognitive strategy (Fernandez-Ruiz et al., 2011).
Participants’ performance on the task can be quantified in terms of the slope of their reward curve relating expected reward to points allocated per hit (Fig. 2.8). The magnitude of the slope corresponds to a measure of how many hits/misses subjects had during the harvest phase. For large positive slopes, performance was good (far more hits than misses) and so a deviation from the Hit$_{100}$Miss$_{0}$ reward schedule results in a greater loss of points relative to the same deviation from optimal that a participant with a smaller slope would experience. Whereas a large negative slope suggests that performance was poor (far more misses than hits) and participants would lose a relatively large number of points from deviating from a Hit$_{60}$Miss$_{40}$ reward schedule compared to a participant with a slightly negative slope.

An implication of this property of the slopes is that it may be harder for mediocre performers (slopes near zero) to determine the optimal reward schedule when compared to either poor or great learners. On the other hand, when the slope is near zero, the choice of reward schedule, by definition, makes little difference in terms of the reward obtained. For the medium and long horizons the average slope was near zero (Fig. 2.8, horizontal dashed lines), and thus deviating from an optimal slope would not have a large impact on reward. Thus, participants may have more difficulty discerning the optimal schedule since they cannot use expected reward as a reliable indicator of what reward schedule to select. The short horizon had a large negative slope, on average, suggesting that the optimal reward schedule may be easier to discern for participants. This is supported by the greater variance of reward schedule selections in the medium and long horizons when compared to the short horizon.
We believe this is the first study to examine whether people are capable of making optimal decisions based on their motor performance, before they have fully mastered the task for which they are making decisions. Other studies have looked at decision making after providing participants with knowledge of their own motor variability on the specific task they will perform, thereby giving them a rich source of information on which to base their decisions on (Trommershäuser et al., 2008). Our study illustrates that individuals can extract information from their learning curve in order to predict future performance, and use that information to make sensible decisions. The optimality of subjects’ decisions depends in part on how well they can extract information from the initial learning or exploration period.

Studies outside the motor control field have shown people exhibit a bias towards risk-aversion and poor statistical decision making (Kahneman & Tversky, 1979). The current study found that, in making high-level decisions about a motor task, participants were risk averse and, with the exception of the short horizon, deviated significantly from optimal when selecting reward schedules. One might speculate that decision making in motor learning tasks may be less prone to such biases because the motor system provides an unbiased source of information that participants can use for their decisions. However, it was up to participants to remember and successfully analyze their motor performance in order to reach an optimal solution. Lapses in attention during initial learning and over/under confidence in learning performance could contribute to difficulty in converting motor performance into the selection of an optimal reward schedule. Defining what “risky” behavior is as a general rule is difficult using this paradigm, as a risky reward schedule for one participant may be the optimal one for another. Whether or not
the decisions participants made correlate with the results of behavioral economic studies is an area for follow-up.
Chapter 3

Reward optimization during motor adaptation

3.1 Abstract
Previous work has shown that, when rapidly reaching towards a screen in which there are both reward and nearby penalty regions, people are able to take into account their movement variability in order to optimize reward. Here we asked how well people can take into account their reach movement variability during a motor learning task, where variability tends to decrease with practice. Specifically, we investigated whether people are capable of taking into account their current and future performance when adapting to a novel sensorimotor transformation, and use that information to make optimal decisions related to task performance. Participants were asked to produce a ballistic movement of a cursor towards a target and were rewarded with monetary compensation each time they successfully reached it. On every fifth trial, participants were given the opportunity to select the size of the target region applied to the next group of five trials, where smaller targets earned participants greater reward. Optimal performance in this task requires that participants accurately assess their performance over the upcoming block of trials. On average, we found that participants selected target sizes that were significantly worse than optimal. Overall, participants’ decisions appeared to take into account both the error they experienced on previous trials and the error in future trials. These results suggest that whereas participants were able to incorporate knowledge of their own learning when selecting targets, they did not do so optimally.
3.2 Introduction

When humans engage in goal directed movement, the presence of sensory and motor noise leads to inherent uncertainty in the outcome of the movement (Harris & Wolpert, 1998). For example, when executing a rapid pointing movement to a target presented on a touchscreen, or when throwing a dart towards the bullseye on a dartboard, there is a substantial amount of variability in the final position of the fingertip or dart. Current motor theory (i.e. optimal feedback control) posits that the sensorimotor system attempts to minimize task-relevant variability by responding more strongly to errors that threaten the successful completion of our goal (Scott, 2004; Todorov & Jordan, 2002). For example, when reaching to grasp a horizontal bar, and given enough time to respond to an error in hand position, the motor system will react strongly to vertical but not horizontal errors in hand position, as the former but not the latter could cause us to miss the bar (Nashed et al., 2012). However, it is often the case that we have limited opportunities to correct for our movement error – such as when throwing a dart or reaching rapidly or without visual feedback of the hand – such that our motor performance has substantial variability.

Factoring in such motor variability may be crucial to the optimal performance of the motor system. In two experiments Trommershäuser and colleagues (2003a; 2003b) showed how people can rapidly estimate their motor variability, and incorporate that information to make optimal decisions about goal directed movements. In these experiments, participants were presented with overlapping reward and penalty regions on a computer touchscreen. Participants received money if they successfully touched the reward region, but lost money if they touched the penalty and/or overlapping regions. A
strict reaction time requirement and the small size of the reward region meant that, given
movement variability, participants could not guarantee earning reward on a given trial.
Given the constraints of the task, for participants to maximize the amount of money they
earn they must factor in their motor variability and the payoff of the reward and penalty
regions. The authors found that participants selected aim points that did not deviate
significantly from optimal, earning them the maximum amount of reward given their
motor variability. This was surprising given that participants were never given explicit
information on the probability of successfully carrying out a potential movement plan.

In a later experiment, Trommershäuser and colleagues (2006) showed that participants
could take their movement variability into account to make decisions about which area of
the screen, containing different reward and penalty regions, to reach towards.
Specifically, participants were required to make a speeded decision between two
alternative penalty/reward configurations similar to those used in the previous studies.
These configurations differed in the distance between penalty & reward regions as well
as the magnitude of the penalty and reward participants could earn. Participants were
informed that they would receive reward for a chosen configuration based on a
simulation of their performance taken from a previous session. On the majority of trials,
participants selected the configuration that would offer them the greatest possible reward
given their motor variability. This result suggests that people might have access to
information regarding their motor variability during explicit decision-making.

While the previous findings documented how people make optimal movements with
fixed motor variability, there is evidence that people can optimally adjust their
movements in response to changes to their motor variability. In motor skill learning,
variability of movements tends to change over time. Compensating for such a change might allow for a more optimal movement plan. Using a similar paradigm as in their earlier studies, Trommershàuser and colleagues (2005) had participants reach to overlapping penalty and reward regions while artificially increasing the variability of their movements. Movement endpoints of participants were varied by a random Gaussian process, which could not be compensated for on a given trial. This perturbation altered where participants should aim to maximize reward. After being introduced to this perturbation, participants took less than 120 trials on average to compensate for the increase in variability and begin making optimal movements to the target. Further evidence comes from a study where children with and without dystonia were able to adjust their movements optimally in response to an imposed decrease and increase in movement variability respectively (Chu, Sternad, & Sanger, 2013).

When learning a novel sensorimotor task, performance – including variability – changes over time. In order to make optimal decisions during learning participants will need access to information about their motor variability in advance of movement planning, and will need to alter their decisions in response to changes in motor variability due to learning. The aim of the current study is to determine whether participants can optimize reward during learning by factoring in their current and future motor performance. In the experiment, participants learned a novel sensorimotor task, and during breaks in learning, were asked to make decisions that affect reward and required knowledge of their current and future motor performance.

The task required participants to apply forces to a rigid joystick instrumented with a force-sensor in order to produce a ballistic movement of a cursor on the screen.
Successful control over the movement of the cursor required participants to learn the mapping between the force they apply to the joystick and the movement of the cursor on the screen. The goal was to move the cursor onto a circular target region in order to receive points, which translated into money at the end of the experiment. On every fifth trial participants were given the opportunity to select a target size—from among five alternatives—to be applied to the next set of five trials they will perform. Target size was inversely proportional to the number of points received for successfully hitting the target. The initial target was the largest target available and participants were informed that they could not return to larger targets; that is, they had to choose whether to stick with the current target or select a smaller one from those available.

We predicted that participants would be able to evaluate their learning parameters (i.e. rate and variability) and make use of that information to select optimal target sizes, thereby maximizing the reward they received during the experiment. Specifically, we predicted that participants would take into account their motor variability on previous trials, and predict motor variability on future trials, in an effort to select the best target size.

### 3.3 Methods

#### 3.3.1 Participants

Nine participants (7 female) were initially recruited as pilot subjects from the Queen’s University community with ages between 18-26 years old ($M = 20.78, SD = 2.39$). An additional eight participants (3 female) between the ages of 18-26 years old ($M = 21.38, SD = 2.77$) were recruited to take part in the main study. All participants received a base pay of $10 / hour, and participants in the main study could earn more by receiving points
during the experiment ($0.02 for every 1 point earned). Participants provided written informed consent, and after the conclusion of the experiment they were debriefed. The experiment was approved by the Queen’s General Research Ethics Board and complied with the Declaration of Helsinki.

3.3.2 Apparatus

Participants used a precision grip to grasp a spherical knob, instrumented with a force sensor (ATI Industrial Automation, N.C.), with their dominant hand. Participants applied isometric force to the force sensor to effectively ‘throw’ the cursor on the screen (Fig. 3.1). The mapping between force direction and cursor direction was the same as a standard computer mouse.

![Diagram](image)

Figure 3.1: Apparatus used for the experiment. Subjects had to push on the force sensor in the direction of the monitor in order to produce a movement of the cursor.

Forces were recorded by the sensor at a rate of 500 Hz in both X and Y directions.
The cursor was modelled as a damped mass:

\[ F = ma + bv \]

where \( F \) is the force applied to the cursor, \( m \) is the simulated mass of the cursor (0.010 kg), \( a \) is the acceleration of the cursor, \( b \) is the viscosity of the cursor (0.035 N/m/s), and \( v \) is the velocity of the cursor. Critically, the force could only be applied to the cursor if the cursor was within 100 mm of the start location. Thus, controlling the cursor involved feedforward or predictive control and participants had to learn the correct impulse (force acting over time) to apply to the joystick in order to produce an accurate cursor movement.

All stimuli were presented onto a vertical screen via rear-projection (NEC UM-330W). Participants made target selections using an external numeric keypad located to the right of the force sensor.

### 3.3.3 Pilot Study

In order to both determine the optimal target sizes to use in the experiment and assess task difficulty we ran an initial group of nine participants. These participants executed reaches towards a small target (10 mm radius) on each trial. Participants did not receive points for hitting the target area, but instead, were given trial-by-trial feedback as to whether they “hit” the target. A hit is defined as any overlap between the area of the cursor and the target circle, plus a 1 mm threshold to account for the fact that objects drawn to the screen 1 mm apart may appear to overlap even when they do not. When participants hit the target, they received feedback in the form of an audible tone (1000 Hz, 50 ms) and the target circle turned white until the end of the trial. Pilot participants performed a total of 400 trials. To encourage good performance, for each time
participants hit the target they received an entry to a raffle for a $20 gift card, such that the more times they hit the target, the greater the probability they had of winning the gift card.

Pilot error data from the first 200 trials was then submitted to a computer algorithm to determine the optimal target sizes to use in the main study. We selected six targets subject to the constraint that participants could only transition from larger to smaller target sizes. The program minimized the amount of time spent on any single target—thereby maximizing the number of switches between targets—while also maintaining an overall hit rate of above 50%. This algorithm provided us with target sizes that would encourage participants to transition between all six targets based on the average performance during the pilot study.

3.3.4 Stimuli
At the beginning of every fifth trial participants viewed a selection of circular targets at the top of the screen (filled blue circle [RGB: 0 100 255], varying radii) with a central black dot (filled black circle, 5 mm radius) in each circle. These targets represented the possible range of target sizes a participant could select from, to be applied to the next five trials. The six target sizes were 32, 38, 48, 64, 104, and 208 mm radius circles. The central dot on each target was the same size as the cursor (filled cyan circle [RGB: 0 255 255], 5 mm radius) and was used to give subjects a central location to aim for on each target (Fig. 3.2). Once the target for the next five trials was selected the target appeared in the center of the screen with the cursor positioned at the start location (filled gray circle [RGB: 109 120 117], 20 mm radius) 300 mm below of the center of the target.
Figure 3.2: Display during every 5th trial. Subjects view available targets to select from on the top of the screen that they can scroll through using the left and right arrow keys on a provided keypad, with the current target highlighted. Subjects confirm their target selection by pressing the down arrow on the keypad.

3.3.5 Procedure

Participants were informed that they would be paid money corresponding to the number of points they earned during the experiment and were informed how the target selection process would occur. Participants earned points corresponding to the size of the target they chose. The largest target (208 mm) was worth a total of 1 point ($0.02 equivalent), with each progressively smaller target being worth one more point such that the smallest target (32 mm) was worth 6 points ($0.12 equivalent). At the beginning of every fifth trial, participants could see the available targets at the top of the screen and scroll through
these targets by pressing the left and right arrow keys on the keypad. The currently selected target was to be highlighted (turned solid white) and displayed in the target location (Fig. 3.2). This allowed participants to see exactly how large the target would appear in the location they would be aiming the cursor towards.

Once participants were satisfied with the target size they were currently selecting, they could confirm the selection by pressing the down arrow on the keypad. This locked in their target choice, and they would proceed to perform five trials of cursor movements aimed towards the target size they selected. Once the five trial period was over, participants could again select a new target size. All participants performed 200 trials and made 40 target decisions (200 trials with 5 trials per target selection).

We enforced two rules on target selection. One, participants could not select a larger target size than the one currently selected. For example, if a subject on the second block selected the 104 mm target size, they could not select the 208 mm target on the third block or any subsequent block. We applied this rule in order to minimize the decision space we would have to optimize. Without this rule, the total number of permutations of target decisions (2.76x10^9) becomes combinatorically explosive. Rather than pruning possible solutions from the tree of target choices, we intended to restrict the target selections using the experimental design. Two, all participants experienced the largest target (208 mm radius) on the very first block. As participants had no previous exposure to the task, we wanted to prevent them from selecting a small target size on the first block and not being able to return to a large target size due to rule one.

At the beginning of each movement trial, the target, start position, and cursor appeared on the screen. Whenever the participant was ready, they applied force on the joystick in
an attempt to land the cursor on the target. Participants could view the cursor as it moved towards the target until it came to a stop. The criteria for hitting the target were the same as the pilot study and explained to participants beforehand. If participants hit the target they received auditory feedback in the form of a beep. Participants were provided with visual feedback at the end of every trial. Text on the screen gave subjects the amount of points they earned on that trial (zero in the case of a miss) along with their running point average. The average gave participants information as to how many points they earned thus far in the experiment, but also provided a good indication of whether their performance was improving.

3.4 Results

3.4.1 Learning Analysis

In general, participants had less variability in the horizontal (x) direction of movement when compared to the vertical (y) direction. Over the course of performing 200 trials, participants reduced the variability in their movements. In the first 20 trials, participants on average had a large standard deviation of movement in the y-axis ($M = 106.88$ mm), with a much smaller variability in the x-axis ($M = 28.89$ mm). In the last 20 trials, participants on average greatly reduced their variability in the y-axis ($M = 43.91$ mm), while reducing variability a smaller amount in the x-axis ($M = 17.17$ mm).

To quantitatively assess performance we computed the movement error: the distance between the position of the cursor when it came to a stop and the center of the target. This was computed as the square root of the sum of the squared horizontal (x) and vertical (y) errors. The four participants shown in Figure 3.3 illustrate the range of performance we observed. All participants had large initial movement errors that then
decreased across trials. The black lines show exponential fits to the data. A one-way repeated measures ANOVA revealed significant main effect of trial block (40 blocks of 5 trials), $F(29,273) = 12.86, p < .001$, on movement error. In addition, a paired t-test revealed a significant difference between movement error in the first block (M = 135.37 mm, SE = 17.04) and last block (M = 35.38 mm, SE = 4.20), $t(7) = 6.003, p = .001$, indicating that on average participants improved performance during the task.

![Learning curves for four different participants. Trials are color-coded by success (green = hit, red = miss). Dashed black line indicates smallest error required for a hit. Solid black line indicates fit of movement error to an exponential function of the form $ae^{bx} + c$, with equation provided. Fits were determined using MATLAB Curve Fitting Toolbox.](image)

Although one might assume that participants always aimed toward the center of the target, no matter the size, it is possible that they allowed greater errors for larger targets. To assess this possibility, we compared movement errors (over the first 200 trials) from
experimental participants to movement errors from participants in our pilot study in which the target was always very small. A two-way group (experimental versus pilot) by trial block mixed factor ANOVA failed to reveal a significant interaction between block and group, $F(39,585) = .759, p = .856$, which suggests that participants in the experimental group aimed to the center of the targets.

Finally, we can determine if participants could hit the smallest target on average by the end of the experiment by comparing the mean error across the last 5 trials of the experiment to 38 mm (radius of smallest target + cursor radius + threshold) using a one-sample t-test. There was no significant difference, $t(7) = -.625, p = .552$, indicating that on average, participants were able to control the cursor accurately enough to enable them to hit even the smallest target by the end of the experiment. This suggests that for the average participant, optimal performance may include selecting the 32 mm target at some point in the experiment.

### 3.4.2 Optimality Analysis

We first determined the optimal target selection for each participant using a brute force approach that computed the number of points the participant would receive, based on their actual performance, for all possible target selection choices (approximately one million) subject to the constraint that a target larger than the currently selected target could not be selected (as in the experiment). The eight panels of Figure 3.4 show, for each participant, movement error as a function of trial. The thick dotted line in each panel shows the optimal target selection for each participant.
In terms of strategy during target selection, it is evident that individual participants tended to either lead or lag behind the transition points of the optimal performance model. Participants P1, P4, P6 and P8 showed a clear pattern of lagging behind the optimal model in their transition points and sometimes failed to select a target selected by the model (see P6 who never selects the smallest target). In contrast, participants P2 and P7 showed the opposite trend of transitioning before the optimal model. A clear dichotomy between lag and lead behavior was less clear in participant P5 who initially lead by transitioning from the 208 mm target directly to the 64 mm target, but then proceeded to lag by not transitioning down to the smallest 32 mm target.

A paired-samples t-test comparing the optimal number of points each subject could have received ($M = 618, SD = 175$) against the actual number of points earned ($M = 563, SD = 149$) revealed that participants had a significantly lower point total than optimal, with an average difference of 55 points (equivalent to $1.10), t(7) = -2.65, p = .033$. We also determined the percentile rank of the target decisions each participant made among all possible decision permutations. On average, participants scored in the 83rd percentile of all decisions ($SD = 18$).
P5: Actual=520 Optimal=559 Local Optimal=557

Movement Error / Target Radius (mm)

Target Radiuses (mm)

- optimal
- local optimal (Lead 5, Lag 0)
- actual

P6: Actual=542 Optimal=587 Local Optimal=586

Movement Error / Target Radius (mm)

Trial
Figure 3.4: Target choices predicted by both an optimal and local optimal model compared to target choices made by the subject. Target choices are represented by colored lines overlaid on a plot of that subject’s movement error (indicated by the black dots). Target size selected is indicated by color (light blue = 208 mm, orange = 104 mm, red = 64 mm, purple = 48 mm, green = 38 mm, navy blue = 32 mm). Number of previous (lag) and future (lead) trials analyzed for the local optimal model is displayed. Amount of points earned by each model is indicated in the graph title.

3.4.3 Lead/Lag Modelling

Whereas the optimal model looks at all 200 trials at once in order to find the best target selections, an arguably more biological model would be constrained to only consider past and (predicted) future trials at each decision point (i.e., every 5 trials). We refer to this model as the ‘local optimal’ model. The model was constrained such that the number of previous trials (lags) and the number of future trials (leads) were constants, with the exception where these numbers were limited by the available number of previous or future trials, respectively. The thin dashed lines in Fig. 3.4 show, for each participant, the predicted target choices for the local optimal model (fit separately to each participant).

The numbers of previous and future trials used in the model is shown for each participant. On average, the model considered 14 previous trials (SE = 4.8) and 18 future trials (SE = 6.5).

To find the local optimum model that optimized reward, we iterated through all possible leads/lags across decision points. Perhaps not surprisingly, a paired-samples t-test revealed that the local optimal model ($M = 615$, $SD = 175$) earned significantly fewer points than the optimal model ($M = 618$, $SD = 175$), $t(7) = -3.51$, $p = .010$. A paired-samples t-test also revealed that the number of points predicted by the local optimal model was significantly greater than the actual number of points earned by the
participants, with an average difference of 51 points (equivalent to $1.02), t(7) = -2.41, p = .047.

3.4.4 Lead/Lag Participant Fitting

To further assess the extent to which participants took into account past and future trials in target choices, we determined the local lead/lag model that best predicted the target selections that participants actually made. Specifically, for each participant, we found the combination of leads and lags—used by the model to maximum reward—that minimized the sum of the squared differences (over all 200 trials) between the target radii selected for the model and the target radii selected by the participant. For comparison, we also determined the best models—in terms of predicting participant target choices—that only considered leads (lead model) or only considered lags (lag model). If participants were indeed both looking ahead and behind when making target selections, we would predict that the combined lead-lag model should result in a smaller square error (i.e. reflect participant choices more accurately) when compared to the fits produced by the lead or lag model. To compare model fits between these three models we took the squared error value for each model fit, and used an $F$ test to compare them.

We first compared the models when combining all participants while allowing separate lead and/or lag parameters for each participant. Thus, there was a total of 8 parameters for the lead model and the lag model (1 model parameter X 8 subjects), and 16 parameters for the combined lead-lag model (2 model parameters X 8 subjects). The combined lead-lag model provided a significantly better fit of the participant data than either the lead model ($F_{(8,304)} = 19.35, p < 0.001$) or the lag model ($F_{(8,304)} = 71.34, p < 0.001$). When the models were compared for each participant separately, we found that
for 7 out of 8 participants the combined lead-lag model was significantly better than the lead model ($p < 0.05$), and that for 6 of 8 subjects the lead-lag model was significantly better than the Lag model alone ($p < 0.05$).

The 8 panels in Figure 3.5 show each participant’s target selections as a function of trials as well as their movement errors. The targets selected by the best-fit lead-lag, lead, and lag models are superimposed. Overall, the lead-lag model appeared to provide a reasonable fit for most participants. The lead-lag model indicated that, on average, participants looked ahead 39 trials ($SD = 47$) and behind 46 trials ($SD = 37$) when determining target selections. All together, this evidence suggests that participants took into account both past errors and predicted future errors when making target selections during learning.
Figure 3.5: Target choices predicted by a lead, lag, and combined lead-lag model, compared to target choices made by the subject. Target choices are represented by colored lines overlaid on a plot of that subject’s movement error (indicated by the black dots). Target size selected is indicated by color (light blue = 208 mm, orange = 104 mm, red = 64 mm, purple = 48 mm, green = 38 mm, navy blue = 32 mm). Number of points earned for the choices of each model indicated in parentheses.

3.5 Discussion

In this study, participants learned a novel motor task that required them to move a cursor to a target and were given the opportunity to periodically select the size of the target, which was inversely related to the reward obtained for a successful hit. The goal of the study was to assess the extent to which participants could take into account future, and generally improved performance, to select target sizes that would maximize reward. We predicted that participants would not base their choices solely on current (or recent) performance but would also consider future performance when selecting a target.

We found that all participants improved on the task in that their reach errors, relative to the centers of the targets, decreased across trials. Whereas a couple of participants reduced errors very quickly, for the most part rather gradual learning curves were observed. Consequently, optimizing reward involved selecting a number of targets of decreasing size across trials. We found that participants’ target choices were on average less optimal than the choices of an optimal model that simultaneously considered performance over all 200 trials. Whereas some participants tended to lag behind the optimal model in terms of switching to smaller targets, others tended to switch earlier than the optimal model. It is possible that participants who lagged behind the optimal model were risk adverse, delaying transitioning to a smaller, more valuable target size out of concern that they would not be able to consistently hit the target on future trials. The
fact that participants could not ‘go back’ and select larger targets (e.g. going from a 32 mm to a 48 mm target) may have encouraged this risk-adverse behavior in these participants.

To assess the extent to which participants considered past and (predicted) future performance when selecting targets, we developed a model that selected targets based on past (lag) and future (lead) trials from the decision points and then found, for each participant, the combination of lead and lags that most closely matched the participant’s target choices. Overall, this lead-lag model provided a reasonable fit to participants’ choices and indicated that they took into account future performance, in addition to past performance, when selecting targets. Whether or not participants may have explicitly considered future performance is not clear.

Our use of models assumes that there is no relationship between target selection and performance. That is, we assume that participants always try to minimize error to the same extent regardless of the size of the target currently selected. Importantly, we did not observe any differences in movement errors across trials between participants in the main experiment (who experience targets of varying size) and participants in our pilot experiment who were presented with a very small target throughout. This suggests that our assumption was valid. However, there is evidence that practicing a task at an optimal difficulty improves performance (Guadagnoli & Lee, 2004) and so it is possible that movement error could be influenced by a participant’s target choice. For example, by selecting a smaller target early on in learning, subjects may compensate by learning the task faster in order to hit the target. On the other hand, given both the novelty and difficulty of the task employed in the current study, participants may have been highly
motivated to reduce errors regardless of target size. Moreover, they may have been highly motivated to improve performance in general as this would enable them to select smaller, more valuable targets.

It has been shown that as choices become more discriminable to a participant (i.e., the difference in expected value between options increases), the proportion of optimal choices increases as well (Jarvstad et al., 2013). In contrast, when there are small expected differences in reward, participants express lower optimality and higher variability in their motor decisions (Battaglia & Schrater, 2007). In the current study, participants may have had difficulty appreciating the differences in expected reward among targets, if so, this may account for why they had difficulty making optimal choices.

While selecting the targets in decreasing order may be optimal, participants on average gained only about 15% of their total reward (calculated from unpublished cumulative reward plots) from performing the first 50 trials of the experiment, where the bulk of the learning took place. That is, participants earned most of their reward once they had completed most of their learning and began collecting reward from the highest value targets. Therefore, there could be a large cost to not switching to the smallest target when this target is the best choice (Subject P8’s performance demonstrates this cost) and we might have expected a bias towards selecting a smaller target early on. In future studies, we might want to focus in on the decisions made during learning, and do so by increasing the expected differences in reward between decisions during this critical period.

Participants’ inability to return to a larger target size than the one they have currently selected may explain why some subjects tended to be risk-averse with their target
choices. That is, selecting a smaller target limits the number of targets that can be selected from in subsequent trials. In previous studies on motor decision making (e.g. Trommershäuser et al., 2005), all choices were available to subjects throughout the experiment and so there was no long-term cost to switching between choices as the experiment progressed. In future studies, keeping the number of choices constant throughout learning would better mimic existing paradigms.
Chapter 4

General Discussion

4.1 Summary of Findings
This thesis investigated critical aspects of decision making in connection with motor learning. We examined whether people can monitor their current performance and predict future performance, in order to arrive at an optimal decision related to reward. Both experiments examined whether participants could maximize the reward they received through prediction of future motor performance.

The first study determined that participants were able to accurately predict their future performance over a short learning duration and use that prediction to select a reward schedule that did not significantly differ from optimal. Over a medium and long length duration of subsequent learning, participants selected reward schedules that were closer to optimal than the ‘average’ reward schedule. This result suggests that participants were able to incorporate information from the initial exploration phase of learning and use that information to select a reasonable reward schedule.

In the second study we compared the target selections made by subjects throughout the course of learning to the target selections made by various models. Participant performance was on average worse than an optimal performer, but was better fit by a model that looked at past and future performance, locally, at each decision point. The results of this study suggest that participants selected targets by both taking into account previously encountered errors as well as successfully predicting errors on future trials.
The results of both studies suggest that people are capable of incorporating knowledge of learning into making reward-related decisions for a skilled motor task. The results of the second study suggest that participants take into account previously experienced error, as well as prediction of future error, when making these decisions.

4.2 Models of Motor Learning
While the optimal performance models used in our experiments do not speculate on the specific learning algorithm participants used, they for the most part assume that participants have perfect information about the errors they will encounter on subsequent trials. Algorithmic models of motor learning that assume a linear or proportional relationship between adaptation and error size have no built-in mechanism that would account for this projection of error beyond the next trial to be performed (Scheidt, Dingwell, & Mussa-Ivaldi, 2001; Thoroughman & Shadmehr, 2000; Wei & Körding, 2009). This discrepancy could be accounted for by the fact that our models may represent more of an explicit contribution of motor learning which is not accounted for by these algorithmic models. Ultimately, the results of these experiments are ambivalent about the specific learning algorithm that participants used. It is possible that participants’ decision making behaviour represents a more general understanding of their learning curve, rather than a specific prediction of error on future trials.

Future attempts at modeling decision making should factor in both variability and uncertainty in participant performance. Our models weighted all movement errors equally, even though some of the more extreme errors may not accurately reflect the participant’s task competence. Real world decision making should be robust to large
outliers as these errors do not accurately represent the current level of motor performance (Körding & Wolpert, 2004).

Additionally, variability in movement becomes a factor when large forces are required for movement, as this increases the level of motor noise (Wolpert, 2007). For this reason, taking into account the amount of force required to produce the movement in both experiments might give us a better model of the decisions participants made. It is interesting that the consequence of larger motor noise may be consistent with behaviour predicted by prospect theory, in that when variability increases, participants may be willing to accept a lower mean reward as long as the variance of reward is minimized (Kahneman & Tversky, 1979).

4.3 Future Directions
Both of these experiments established explicit reward for movement outcomes. However, in real world decision making there is often uncertainty in the amount of reward that can be earned. For example, even if a tennis player is able to land the ball in the location they desired, they have no guarantee that their opponent will not be able to successfully return it. As a direct follow-up, it may be beneficial to model participant choices in the face of uncertain reward.

In future work, we hope to address the topic of investment in learning. In our daily lives we often must decide whether to invest in learning a new skill, ranging from a programming language to a new tennis grip, based on an evaluation of the costs and rewards associated with learning. Making an optimal decision to invest in learning depends on being able to correctly evaluate the costs of performance (e.g. risk of failure, loss of time/money), and weigh them against the potential rewards (e.g. happiness,
productivity, social benefits). Furthermore, with tasks that require longer investment times, energetic and biomechanical costs may become a factor in a model of decision making (Cos et al., 2014). Successful evaluation of our learning parameters (e.g. rate of learning, motor variability) and knowledge of the frequency or duration over which we can make use of our learned ability can further allow us to determine whether an investment in motor learning is worthwhile.
References


