Internal Models of Embodied Dynamics: A Computational Theory of Learning in Routine Interactive Behavior

by
Christopher Robert Sims

A Thesis Submitted to the Graduate Faculty of Rensselaer Polytechnic Institute in Partial Fulfillment of the Requirements for the degree of DOCTOR OF PHILOSOPHY

Major Subject: Cognitive Science

Approved by the Examining Committee:

_________________________________________
Dr. Wayne D. Gray, Thesis Adviser

_________________________________________
Dr. Michael Schoelles, Member

_________________________________________
Dr. Brett R. Fajen, Member

_________________________________________
Dr. Qiang Ji, Member

Rensselaer Polytechnic Institute
Troy, New York
May, 2009
CONTENTS

Internal Models of Embodied Dynamics: A Computational Theory of Learning in Routine Interactive Behavior ................................................................. i

LIST OF TABLES ......................................................................................................................... vi

LIST OF FIGURES ........................................................................................................................ vii

ABSTRACT ........................................................................................................................................... xiii

1. Introduction ...................................................................................................................................... 1
   1.1 Why study routine behavior? ........................................................................................................ 6
   1.2 Organization of the thesis .............................................................................................................. 7

2. Historical review ........................................................................................................................... 10
   2.1 Studies of routine interactive behavior ........................................................................................ 10
       2.1.1 Summary of previous findings ............................................................................................... 15
   2.2 Research on internal models in sensorimotor control ................................................................. 16
   2.3 Summary ...................................................................................................................................... 19

3. Learning and exploiting internal models of embodied dynamics ................................................. 20
   3.1 Learning forward models .............................................................................................................. 21
   3.2 Learning inverse models .............................................................................................................. 25
       3.2.1 Direct inverse modeling ........................................................................................................ 26
       3.2.2 Distal supervised learning ................................................................................................... 28
       3.2.3 Trajectory sampling ............................................................................................................. 32
   3.3 Summary of learning algorithms .................................................................................................. 33
   3.4 Exploiting internal models for embodied cognition ...................................................................... 35
       3.4.1 The adaptive internal model framework .............................................................................. 35
       3.4.2 Biological evidence for the adaptive internal model framework ........................................ 38
       3.4.3 Experimental data set ........................................................................................................... 41
       3.4.4 Cognitive model based on the adaptive internal model framework ................................... 43
   3.5 Summary ...................................................................................................................................... 47

4. Bayes’ Ball experiment 1 .............................................................................................................. 49
4.1 Hypotheses ........................................................................................................... 52
4.2 Method ................................................................................................................ 53
  4.2.1 Participants ...................................................................................................... 53
  4.2.2 Apparatus and materials ................................................................................ 53
  4.2.3 Design ............................................................................................................ 56
  4.2.4 Procedure ....................................................................................................... 57
4.3 Results .................................................................................................................. 57
  4.3.1 Performance during practice trials of the experiment............................... 57
  4.3.2 Performance during test trials of the experiment ........................................ 59
  4.3.3 Post-experiment questionnaire ..................................................................... 65
  4.3.4 Summary and discussion of results ............................................................... 65
4.4 Ideal performer analysis ..................................................................................... 67
4.5 Cognitive models of Bayes’ Ball ..................................................................... 75
  4.5.1 Time estimation and constraints on performance ........................................ 76
  4.5.2 Cognitive model based on the adaptive internal model framework .......... 77
  4.5.3 Reinforcement learning model .................................................................... 87
4.6 Summary ............................................................................................................. 92
5. Bayes’ Ball experiment 2 .................................................................................... 94
5.1 Overview and hypotheses ................................................................................ 94
5.2 Method ................................................................................................................. 95
  5.2.1 Participants ................................................................................................... 95
  5.2.2 Apparatus and Materials ............................................................................. 96
  5.2.3 Design .......................................................................................................... 97
  5.2.4 Procedure ..................................................................................................... 99
5.3 Results ................................................................................................................. 100
  5.3.1 Mean points per trial .................................................................................. 100
  5.3.2 Evidence for score improvement across blocks ......................................... 102
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.4.11</td>
<td>Velocity profiles for the initial movement</td>
<td>142</td>
</tr>
<tr>
<td>6.4.12</td>
<td>Vertical deflections in movement trajectories</td>
<td>145</td>
</tr>
<tr>
<td>6.4.13</td>
<td>Summary and discussion of results</td>
<td>148</td>
</tr>
<tr>
<td>6.5</td>
<td>Minimum jerk model of the via-point reaching task</td>
<td>149</td>
</tr>
<tr>
<td>6.6</td>
<td>Elements of an ideal performer analysis</td>
<td>155</td>
</tr>
<tr>
<td>6.6.1</td>
<td>Optimal control problem formulation</td>
<td>158</td>
</tr>
<tr>
<td>6.6.2</td>
<td>Optimal control model results</td>
<td>160</td>
</tr>
<tr>
<td>7.</td>
<td>Summary and conclusions</td>
<td>162</td>
</tr>
<tr>
<td>7.1</td>
<td>Introduction</td>
<td>162</td>
</tr>
<tr>
<td>7.2</td>
<td>Summary</td>
<td>163</td>
</tr>
<tr>
<td>7.3</td>
<td>Discussion</td>
<td>168</td>
</tr>
<tr>
<td>7.3.1</td>
<td>Computational models in theory development</td>
<td>168</td>
</tr>
<tr>
<td>7.3.2</td>
<td>Implications for other research</td>
<td>170</td>
</tr>
<tr>
<td>7.4</td>
<td>Conclusions</td>
<td>172</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>173</td>
</tr>
<tr>
<td></td>
<td>Appendix A: Debriefing questionnaire used for experiment 1</td>
<td>184</td>
</tr>
<tr>
<td></td>
<td>Appendix B: Derivation of the ideal performer analysis</td>
<td>185</td>
</tr>
<tr>
<td></td>
<td>Appendix C: Derivation of the modified minimum jerk model</td>
<td>187</td>
</tr>
<tr>
<td></td>
<td>Appendix D: Electronic archive of materials</td>
<td>193</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 4-1. Mean target-relative stopping position (delta) and score as a function of target distance and penalty configuration. .................................................................62

Table 4-2. Mean target-relative stopping position (delta) and score for human subjects and the ideal performer analysis.................................................................73

Table 5-1. Summary of the within-subject manipulations for Bayes' Ball experiment 2. The order of distances was counterbalanced across subjects, while the order of penalty configurations was randomized subject to the constraint that Both-penalty trials occurred first in each sequence of three blocks ...........................................99

Table 5-2. Mean scores for each combination of noise condition, penalty configuration, and target distance. ................................................................................................102

Table 5-3. Mean delta values for each condition. ................................................................................104

Table 6-1. Proportion of correct responses for each task condition .........................................................128

Table 6-2. Proportion of incorrect responses by task condition ............................................................130

Table 6-3. Mean initial movement latency by experimental condition ......................................................132

Table 6-4. Proportion of wrong-way movements by task condition ..........................................................135

Table 6-5. Mean total trial time for each experimental condition .............................................................136

Table 6-6. Mean movement speed by experimental condition .................................................................139

Table 6-7. Interaction between condition and task for movement duration from the waypoint to the response circle. .......................................................................................142
LIST OF FIGURES

Figure 3-1. Diagram of a typical feedforward neural network with two input units, three hidden units, and one output unit. Information flows from left to right through the network. .........................................................................................................................23

Figure 3-2. (Top row) Two fictional data sets used to demonstrate how a neural network can learn a forward model. The x axis represents the model input, and the y axis represents the model output. For each graph, the black points are the noisy training data, while the red curve shows the predictions from the trained forward model. Data on the left was generated from the function \( y = 4(x - 0.5) \), while the data on the right was generated from \( y = x + 0.3\sin(2\pi x) \). (Bottom row) For each data set, the red curve indicates the absolute difference between the model’s output and the underlying (noise-free) function used to generate the training data..........................25

Figure 3-3. The reciprocal relationship between forward and inverse models. x is the current state information, a is an action, and y is a resulting or desired state........26

Figure 3-4. Direct inverse modeling applied to two artificial data sets. (Top row) The input and output values have been reversed, such that the desired system output is plotted on the horizontal axis and the control signal is plotted on the vertical axis. In both cases, the model fails to learn an appropriate control policy. (Bottom row) Control error for the two inverse models. Control error is defined as the absolute difference between the actual and desired outcome resulting from following the learned control policy shown in the top row.................................................................28

Figure 3-5. Using distal supervised learning to train an inverse model. The dashed line indicates the training signal used to adjust the inverse model.................................29

Figure 3-6. Distal supervised learning applied to two artificial data sets. A forward model was first trained on each data set; this model was then used to train an inverse model. For the first data set (left column) the algorithm is able to acquire an accurate control policy. For the second data set (right column) the algorithm fails to learn an appropriate control policy, demonstrated by the large prediction error (bottom right graph)........................................................................................................31
Figure 3-7. Control policies found by sampling from a forward model. For both functions, arbitrarily low error (relative to the error in the forward model) can be achieved by varying the number of actions sampled. In these examples, 100 random actions were sampled for each desired outcome.

Figure 3-8. The adaptive internal model framework.

Figure 3-9. The reaching task studied by Augustyn and Rosenbaum (2005). (a) On each trial, the subject placed a cursor along a line separating two circular targets. (b) The line and one of the targets (randomly chosen) disappeared, and the subject was required move the cursor to the remaining target within a short interval. (c) Fractional distance of the chosen start position $D_L$ from the lower of the two targets, as a function of the targets’ relative widths. Error bars indicate +/- 1 SE. Figure adapted from (Augustyn & Rosenbaum, 2005).

Figure 3-10. Comparison of human subject data (a) to a cognitive model (b) based on acquiring and exploiting an internal model of the embodied dynamics of Fitts’ law.

Figure 3-11. Average learning curve for the cognitive model, defined in terms of the mean square difference between selected and optimal actions.

Figure 4-1. Demonstration of the Weber law of interval timing. The three curves show the distribution of subject-estimated temporal intervals lasting for 8, 12, and 21 seconds. Note that the width of the distribution (its standard deviation) increases with increasing interval magnitude. The plot markers indicate human subject data, while the smooth curves are model predictions based on a psychological theory of time estimation. Figure taken from (Taatgen, Rijn, & Anderson, 2007).

Figure 4-2. Apparatus used in the Bayes’ Ball task. (a) During the practice trials subjects attempted to stop the ball directly on the red ‘x’. (b) During the test trials, subjects attempted to maximize their point score by stopping in the reward region (green) and avoiding the penalty region(s) (red). This example shows the penalty-after configuration. (c) The penalty-before condition, and (d) the both-penalty condition.

Figure 4-3. Summary of experimental design used for Bayes’ Ball experiment 1.
Figure 4-4. Standard deviation in subjects' targeting accuracy (delta) as a function of target distance. Error bars indicate 95% confidence intervals.

Figure 4-5. Mean test trial score across epochs of the experiment. Each epoch consists of 225 trials. Error bars indicate 95% confidence intervals.

Figure 4-6. Mean target-relative stopping error (delta) for each distance and penalty configuration.

Figure 4-7. Compensation relative to mean stopping position on both-penalty trials for each target distance. Error bars indicate 95% confidence intervals.

Figure 4-8. Comparison of performance of a single subject to an ideal performer calibrated to that subject. (Top) The smooth curve shows the predicted optimal stopping point distribution, while the histogram (blue bars) shows the empirically observed distribution. (Bottom) The dashed curve shows the expected value (in points) as a function of mean stopping position. The maximum of the dashed curve corresponds to $U(\tau_{opt})$. The marked point shows the mean stopping position and resulting score for the human subject’s data shown in the histogram at top. The error bars (vertical and horizontal) indicate 95% confidence intervals in mean position ($x$ error bars) and score ($y$ error bars).

Figure 4-9. Compensation relative to mean stopping position on both-penalty trials for each target distance. Solid lines reflect empirical data, while the dashed lines are the ideal performers. Error bars on the human data indicate 95% confidence intervals.

Figure 4-10. Neural network structure employed for the internal model of embodied dynamics learned in the Bayes’ Ball experiment.

Figure 4-11. Learning curves for human subjects and the adaptive internal model architecture across practice blocks of the experiment. The curves represent mean absolute difference between stopping position and the target center. Each block consists of three trials (one trial at each target distance).

Figure 4-12. Internal model of embodied dynamics in the Bayes’ Ball experiment. (a) The predicted mean ball position (black curve) as a function of subjective elapsed time, with 95% confidence intervals overlaid (red curves). The marker points indicate the model experience from previous trials. (b) The predicted standard
deviation in ball position as a function of subjective elapsed time. Although the Weber law predicts a linear increase in standard deviation with..........................83
Figure 4-13. A comparison of mean score across four epochs of the test trials for human subjects (solid markers) and model simulations (open markers).........................84
Figure 4-14. Comparison of human and model performance on the test trials of Bayes’ Ball experiment 1. The top row compares the human data (left graph) with the parameter free model (right graph). The bottom row compares human data (repeated from the top row) with a model incorporating two biases. See text for details. For the bottom two panels, $r^2 = 0.95$.........................................85
Figure 4-15. Comparison of the reinforcement learning, embodied dynamics model, and human performance over blocks of test trials.................................92
Figure 5-1. Sample feedback given to participants after each block of 225 trials.............97
Figure 5-2. Mean scores for each combination of target distance, penalty configuration, and noise condition. (Left) The data for the no-noise condition. (Right) Data for the noise condition. Error bars indicate 95% confidence intervals..........................101
Figure 5-3. Mean delta values for each condition. (Left) Data from the No-noise condition (Right) Data from the Noise condition. Error bars indicate 95% confidence intervals..................................................103
Figure 5-4. Mean delta by trial number and transfer condition. The transfer from block 1 to block 2 occurred after trial 225. Error bars on the penalty-after transfer condition indicate 95% confidence intervals..................................................107
Figure 5-5. Comparison of mean delta empirically observed during the experiment (top row) versus subjects’ estimates of the optimal strategy for the task (bottom row). The left column is data from the no-noise condition, while the right column is data from the noise condition. Error bars indicate 95% confidence intervals.............108
Figure 5-6. Comparison of human subject and ideal performer delta, for each combination of distance, penalty configuration, and noise condition. Human data are presented in the top row, model predictions are in the bottom row. The left column shows data from the no-noise condition, while the data in the right column is from the noise condition.................................................................111
Figure 5-7. (Left) Correlation between predicted optimal and empirical scores. (Right) Correlation between model-fit standard deviation in ball position and empirically observed standard deviation. The diagonal dashed lines indicate the unit slope corresponding to perfect correlation, the solid line gives the best fit linear regression.

Figure 5-8. Expected utility curves for a single subject on the penalty-before configuration at each of the three target distances (distance measured in pixels).

Figure 5-9. Comparison of empirical (top row) and model predictions (bottom row) in terms of delta. The left column is data from the no-noise condition, while the right column is from the noise condition. Error bars on the human data are 95% confidence intervals.

Figure 6-1. (a) Sequence of events on each trial. Subject-triggered events are shaded in gray. (b) Layout of the task interface for the bilateral experiment condition, after the mouse entered the start circle. (c) Task interface for the unilateral experiment condition.

Figure 6-2. Two-way interaction between response mode and task context for the overall percent of correct trials. Error bars indicate 95% confidence intervals.

Figure 6-3. Initial movement latency as a function of context and response mode. Error bars indicate 95% confidence intervals.

Figure 6-4. Empirical histogram of movement latencies for each response mode and block context. The bin size used for the histogram is 25 ms.

Figure 6-5. Mean total trial time over blocks of 20 trials. Error bars indicate 95% confidence intervals.

Figure 6-6. Mean movement speed by context and response mode. Error bars indicate 95% confidence intervals.

Figure 6-7. Distribution of movement speeds for each combination of context and response mode.

Figure 6-8. Velocity profiles for the hand movement from the start circle to the waypoint, by experimental condition. (a) Velocity as a function of percentage of the total movement duration. (b) Temporally aligned mean velocity profiles.
Figure 6-9. Vertical deflection in mouse movements from the start circle to the waypoint. Triangles indicate trials where the response was 'same', circles correspond to responses of 'different'. Filled markers indicate pure blocks, while open markers are data from mixed blocks. Solid lines are from the unilateral condition, dashed lines are from the bilateral condition.

Figure 6-10. Comparison of minimum jerk trajectories (solid lines) and empirical trajectories (dashed lines). (a) Minimum jerk trajectory computed with waypoint passage time optimized to maximize movement smoothness. (b) Minimum jerk trajectory with waypoint passage time set to the empirically observed value. Empirical data is based on the bilateral mixed condition.

Figure 6-11. Comparison of empirical and predicted velocity profiles. (a) Human data, re-plotted from Figure 6-8. (b) Velocity profiles predicted by the minimum jerk model for each experimental condition, with waypoint passage time and total movement time constrained to the empirically observed values.

Figure 6-12. Optimal velocity profiles as a function of (a) percent of total movement duration, and (b) trial time, computed for two values of the minimum decision time parameter.
ABSTRACT

How are human cognitive, perceptual, and motor processes organized and coordinated towards the efficient achievement of goals in routine interactive behavior? Despite the simplicity of the question, its answer is at present poorly understood. The goal of this thesis is to provide a unifying explanation for the intelligence and behavioral richness inherent to routine human activity. This explanation centers on the capacity to acquire and exploit internal models of embodied dynamics. Embodied dynamics are the recurring interactions between cognitive, perceptual, and motor processes with external tasks and environments. Internal models are formal constructs that have largely been studied in low-level sensorimotor control. The emphasis of this research is on how these two elements can be combined to generate novel and surprising predictions regarding the capacities of human performance. This thesis demonstrates, through the convergence of three empirical studies, mathematical optimality analysis, and computational cognitive modeling, the human capacity to acquire and exploit such internal predictive models.
1. Introduction

“For the things we have to learn before we can do them, we learn by doing them” (Aristotle, 350 B.C.E.)

Nearly all of human activity can be described as an intricately coordinated dance of low-level cognitive, perceptual, and motor components, each taking small steps of progress towards a larger goal. The elementary operations performed by these components, unfolding on the timescale of 1/3 of a second, form the *embodiment level* of cognition (Ballard, Hayhoe, Pook, & Rao, 1997)—the level of analysis at which the constraints of the low-level components first have implications for the ability to achieve goals in the world. While the field of cognitive science and its precursors have a long history of studying the basic properties of cognition, perception, and motor control in isolation—for example, the rate of memory decay as a function of time and practice (Ebbinghaus, 1913) or how movement duration scales with magnitude (Fitts, 1954)—the question of coordination and co-adaptation among these components has received considerably less attention.

There may be several reasons why coordinated interactive behavior is poorly understood compared to the performance of cognitive, perceptual, and motor processes in isolation. One reason is simple pragmatism. It might be argued that the basic properties of each system must be understood in sufficient detail before they can be studied in concert. There is reason to suspect however, that this divide-and-conquer approach may not ultimately yield a satisfactory theory of the human mind (Newell, 1973). Others have adopted the independence of processing components as a strong philosophical or theoretical stance, rather than as a matter of expedience in scientific practice (Fodor, 1983).

Another reason for the lack of progress on understanding the integration of cognitive, perceptual, and motor processes is that the same behavior can be studied on multiple levels of analysis. Newell (1990) defined a hierarchy of four timescales of human behavior: the biological, cognitive, rational, and social bands. Within this hierarchy, each band spans three orders of magnitude in terms of the relevant timescale of its operations. Processes lasting 100s of milliseconds define the lower end of the cognitive band, with *unit tasks* lasting on the order of 10 seconds defining the upper
limit of this band. The embodiment level of analysis can be seen as focusing on the lower end of the cognitive band of analysis.

At higher levels of analysis, the coordination of behavior can be studied in terms of disembodied scheduling processes, often modeled after computer operating system metaphors (Kieras, 2007; Salvucci & Taatgen, 2008). On this level of analysis, the coordination of behavior occurs largely or wholly independent of the dynamics of the components that must be coordinated. The danger of this approach is that by abstracting away much of the richness of behavior, what is left to be explained by the theory is after all a non-problem. On a lower level of analysis, the biological band, the problem of coordination disappears, since notions of tasks or goals are ill defined on the neural level, and since physical constraints such as neural conduction delays preclude the possibility of any interaction between cognition, perception, and motor control. Both of these research agendas have made tremendous progress in our understanding of the operation of the human mind. However, a central purpose of this thesis is to demonstrate the productiveness of studying interactive behavior on the embodiment level. On this level of analysis, even routine interactive behavior will be shown to possess an exquisite behavioral richness, in terms of adaptation to and exploitation of the dynamics of the task, cognitive system, and ongoing performance.

The approach I will describe builds on Anderson’s rational analysis framework (J. R. Anderson, 1990), or the idea that “we can understand a lot about human cognition without considering in detail what is inside the human head. Rather, we can look in detail at what is outside the human head and try to determine what would be optimal behavior given the structure of the environment and the goals of the human” (p. 3). Concretely, this involves specifying in formal terms the goal or goals of the cognitive system, the statistical structure of the environment to which the cognitive system is adapted through evolution, and computing the resulting optimal behavior. Limitations, such as due to evolutionary constraints, can also be incorporated into the analysis.

The rational analysis approach stands in contrast to the historically more common approach to explaining human behavior through computational models, where the only justification for the existence of some putative mechanism is its ability to reproduce the empirical data for which it was created. The rational analysis approach has previously
been used to demonstrate that human memory closely mirrors the information demand structure of the natural environment, such that the information that is most likely to be needed is also the least likely to be forgotten (J. R. Anderson & Schooler, 1991). In another context, human perception in many visual illusions has been explained as a rational adaptation based on the sensory statistics of the broader visual environment (Purves, Williams, Nundy, & Lottom, 2004).

The value of the rational analysis approach is that it allows researchers to develop computational models of cognition that explain why human performance is the way it is. Humans forget information at a particular rate because the natural environment suggests that this information is unlikely to be needed. Humans misperceive illusions because these illusions violate the statistical structure of the natural visual world. A central theme in the rational analysis approach is the assumption that components of the cognitive system are adapted over the course of evolution and human development to the statistics of the external environment. These stable interactions between components such as memory or visual perception and broad statistical properties of the environment constitute examples of component-environment interactions, unfolding on a lifetime or evolutionary timescale.

While rational analysis goes beyond the historical tendency of studying components in isolation to instead focus on component-environment interactions, it is often the case that the efficient achievement of goals requires accounting for short-term component-task interactions as well. For example, the motor variability associated with hand movements may have different implications for the effectiveness of a planned motor trajectory in different tasks—consider the act of reaching for a target performed as part of a surgical operation versus grasping a cup of coffee. Different tasks impose varying performance criteria that are obscured by considering only the aggregate statistical properties of the natural environment as a whole. Thus, it is necessary that in planning and controlling behavior, the cognitive system incorporate component-task in addition to component-environment interactions.

Finally, it is also the case that the efficiency of interactive behavior depends on accounting for component-component interactions. Consider using a calculator to sum a list of numbers. The motor system must be coordinated with cognitive and visual
processes in order to efficiently complete the task. In reaching to press the addition button, the hand movement might be slowed to allow the cognitive system time to retrieve the next number to be added from declarative memory (that this sort of cognitive-motor coordination actually occurs will be empirically demonstrated in the third experiment of this thesis). In most tasks, there is little to be gained from a motor system that executes movements as fast as possible—doing so increases the chance of motor error and expends more energy without improving performance on the task. Instead, the ideal control of the motor system depends on the processing dynamics of the non-motor components of ongoing performance. This constitutes an example of the important for accounting for component-component interactions.

Given the vast space of possible interactions between cognition, perception, motor control and changing tasks, it is unlikely that efficient coordination of interactive behavior could be achieved purely through evolutionary means or adaptation to stable statistical properties alone. Instead, the cognitive system must somehow flexibly modulate its behavior based on novel component-task and component-component interactions, a process that requires learning and skill adaptation as well as the evolution and development that are the traditional focus of rational analysis (Oaksford & Chater, 2007). Therefore, the use of rational analysis in this research represents an extension beyond its traditional usage, to include consideration of not just component-environment interactions, but component-task and component-component interactions as well. This generalized approach to understanding rational behavior will be referred to as optimality analysis, or ideal performer analysis in the chapters that follow.

How are embodied cognitive, perceptual, and motor components coordinated in a manner that is both efficient and relevant to the immediate tasks and goals facing the individual? This dissertation will address the question in some detail, by employing both experimental study and computational cognitive modeling. In particular, two hypotheses will be advanced:

1) The coordination of interactive behavior is sensitive to the low-level dynamics and variability intrinsic to human cognitive, perceptual, and motor processes, and to the external task environment.
The cognitive system achieves this coordination in routine behavior by learning and exploiting internal models of these embodied task dynamics.

The word *dynamics* is employed here with the same intent as used in mathematics and physics: it refers to a specified relationship between the change in one quantity or variable and the resulting changes in another. The phrase *embodied task dynamics* is intended to reflect the full combination of component-environment, component-component, and component-task interactions, as all three bear implications for the effectiveness of behavior in a specific setting. Fundamentally, these interactions are located neither in the head nor in the external world, but instead are mutually defined and dependent. Understanding how cognition adapts to such dynamics requires studying both the human agent and the external task environment in tandem.

In treating internal processes such as memory and motor systems on the same footing as external processes such as the sequential and temporal structure of tasks and artifacts, this approach has its origins in the philosophical positions of embodied cognition (Wilson, 2002) and active externalism (Clark, 2003). Active externalism refers to the philosophical belief that the mind and external environment are coupled systems, and that the human reliance on an internal resource such as declarative memory should not be studied or understood as a distinct phenomenon from the reliance on external memory resources, such as notebooks or scraps of paper. In this work, it will similarly be proposed that the mind does not discriminate between internal and external dynamics when acquiring internal models. From the perspective of predictability, regularities and structures in the world are no different than regularities and structures in cognitive operations, perception, or motor control.

The idea of “internal models of the world” has at times been contentious in psychology and cognitive science, and has sparked research programs that attempt to completely eliminate their role in models of cognition, most famously demonstrated by the robotics work of Brooks (1991), and Agre and Chapman’s *PENGI* theory of activity (Agre & Chapman, 1995). However, as others have noted (M. L. Anderson, 2003; Noë, 2005; O'Regan & Noë, 2001), internal models that are grounded through experience in the world are compatible with the embodied cognition approach, and perhaps essential to explaining many aspects of human cognition.
1.1 Why study routine behavior?

In both of the main hypotheses of this proposal, the role of routine, rather than novel or extraordinary human behavior has been emphasized. This reflects the belief that even mundane tasks carried out by humans exhibit a degree of intelligence not captured by the majority of existing theories of cognition. As has previously been noted (Rosenbaum, 2005; Rosenbaum, Carlson, & Gilmore, 2001), it is now possible to construct computer programs that are able to beat the world’s best human chess players, yet robots have not advanced to the point of being able to climb a tree as well as a five year old child. This is not likely due to a limitation in the theoretical understanding of human sensory systems or motor effectors, but rather the efficient control and coordination of these component processes for achieving fluid interactive behavior in natural tasks. Like vision and the ability to communicate linguistically, the ability to control and coordinate interactive behavior appears completely effortless upon introspection. The underlying complexity remains transparent unless it is willfully adopted as the focus of study.

If it is accepted that even routine behavior exhibits a remarkable degree of intelligence, there are two fundamental properties of routine tasks that are essential to explaining the human ability to intelligently execute them. First, the vast majority of routine tasks carried out by humans are benign (Agre, 1988), in the sense that it is rarely catastrophic if mistakes are made. In situations where mistakes can be costly, social institutions intervene during the learning process and restructure the task to ensure that it is made benign. In the case of learning to drive a car this takes the form of legislative restrictions on when and with whom novices can drive, or providing supervised instruction in cars with a second steering wheel or set of pedals. In the case of performing surgery, an active area of current research is the development of training simulators and haptic devices such that the cost of mistakes is eliminated (e.g., O'Malley, Gupta, Gen, & Li, 2006).

Second, by virtue of the routine nature of a task, humans can “amortize the cost of algorithms” (Ballard, 1997, p. 9) over repeated experience, thus greatly simplifying the computational learning burden. Rather than solving an optimal scheduling problem for a task the first time it is encountered, solutions can be incrementally constructed over time. Even when a task is performed for the first time, it is more likely than not that the basic
activities involved in the task are themselves highly practiced (for instance, the use of a computer mouse, encoding and retrieving facts from memory, etc). It is by studying routine tasks that exhibit these two features (benign outcomes and amortized learning costs) that we can best hope to uncover the mechanisms underlying the intricate coordination of interactive behavior. As noted by Carlson, “Most cognitive activity is routine and thus skilled, and must therefore be understood by studying the acquisition of cognitive skill” (1997, p. 12).

1.2 Organization of the thesis

Chapter 2 begins with a review of recent literature on the coordination of cognitive, perceptual, and motor processes in routine interactive tasks. The studies reviewed in this chapter provide examples of the relevant phenomena to be addressed by this dissertation. While the empirical findings have previously been accounted for by a variety of loose verbal theories or explanations (scheduling, metacognition, etc.), a goal of this dissertation is to provide a unifying explanation in terms of adaptation to and exploitation of internal models of embodied dynamics. The concept of an internal model is formally defined, based on its usage in the literature on low-level motor control. The available evidence supporting the existence of internal models in the motor domain is reviewed.

Chapter 3 contains a computational exploration of properties of internal models, with particular regard to how they can be learned through self-directed experience in the world. Alternative algorithms that have previously been proposed as mechanisms for motor learning are evaluated on a simple, abstract learning problem. The results demonstrate that several currently known algorithms possess substantial limitations, even on relatively simple toy problems. A computational framework for learning and exploiting internal models for embodied cognition is then proposed. This framework is called the adaptive internal model framework. The biological basis for the framework is supported by reference to the known anatomical and functional roles of the human cerebellum, basal ganglia, and cerebral cortex. The computational and explanatory sufficiency of the framework is then demonstrated on a simple perceptual motor experiment that has previously been studied (Augustyn & Rosenbaum, 2005).
Chapter 4 introduces a novel paradigm, known as Bayes’ Ball. This paradigm is used to demonstrate that humans are capable of acquiring and exploiting internal models of the variability inherent in their ability to estimate short time intervals. In particular, the Weber law of interval timing (Gibbon, 1977; Staddon & Higa, 1999) states that the standard deviation of an estimated time interval increases linearly with the magnitude of the interval to be estimated. Subjects in the Bayes’ Ball experiment are shown to not only conform to the Weber law, but to demonstrate implicit awareness of the implications of the law in their behavior. An ideal performer analysis shows that human performance in the experiment is very nearly optimal, given the constraints imposed by the task and subjects’ individual abilities to accurately estimate time intervals. A computational model, based on the adaptive internal model framework, is then constructed and analyzed. The model is able to capture detailed properties of human performance, including the time course of learning in the experiment. An alternative model is then constructed, one that avoids the use of internal models, and is shown to be fundamentally at odds with the empirical findings from the experiment.

Chapter 5 presents the results of a second experiment using the Bayes’ Ball paradigm. This experiment was designed to address several questions that arose based on the results from the first study. In addition, a between-subjects manipulation introduced artificial uncertainty into task performance. It is shown that subjects adapted to this additional uncertainty, and altered their performance appropriately. This is given as evidence that learning of embodied dynamics can occur rapidly in novel tasks. Finally, the results of the second study demonstrate that the learning and exploitation of embodied dynamics can occur implicitly, without conscious awareness or deliberate intention to do so.

Chapter 6 introduces an additional paradigm that is designed to test the coordination of cognitive, perceptual, and motor processes in a somewhat more complex task environment, where subjects are given greater freedom in terms of the space of possible task strategies. The particular task requires that subjects coordinate a motor reaching task with a simple object discrimination task. The paradigm exploits the fact that in task-switching scenarios, the difficulty of cognitive processing is increased, leading to slower cognitive processing and increased error rates. The results of the study demonstrate that
motor performance is adaptively altered to incorporate the properties of ongoing cognitive processing.

Chapter 7 concludes the thesis with a summary and discussion of findings.
2. Historical review

The coordination of interactive behavior cannot be achieved by a fixed, task-independent mechanism acquired through evolution or slowly over the course of a lifetime, although undoubtedly both evolution and development play an important role in determining human behavior. Instead, it is argued that routine interactive behavior relies on dynamic adaptation to and exploitation of embodied task dynamics. This chapter reviews recent studies on the coordination of cognitive, perceptual, and motor processes in routine interactive tasks. The studies provide examples of the relevant phenomena to be addressed by this dissertation. While the empirical findings presented here have previously been accounted for by a variety of verbal theories or explanations, a goal of this dissertation is to provide a unifying explanation in terms of adaptation to and exploitation of internal models of embodied dynamics. Finally, the concept of an internal model is more precisely defined, based on its usage in the literature on low-level motor control.

2.1 Studies of routine interactive behavior

In a clever study on the origins and nature of skill acquisition (Agre & Shrager, 1990), a temporary office worker was hired and then surreptitiously videotaped by cameras mounted in the ceiling as she copied papers from a bound volume on a photocopier machine. The subject was not made aware that she was being videotaped or participating in an experiment in order that her behavior would be no different than in a typical office setting. The entire experiment consisted of copying a 17-page paper from a 296 page spiral-bound book, and as the primary behavioral measure the investigators analyzed the time required to photocopy successive pairs of pages, for example, the time between copying pages 1–2 versus the time between copying pages 16–17. The outcome of the experiment demonstrated a systematic speedup in the subject’s time to copy successive pairs of pages. The shape of the speedup could be closely fit by a power function, providing further evidence for the ubiquitous “power law of practice” even in the behavior of a single individual.

While numerous explanations for the power law of practice have been offered, for example production compilation (J. R. Anderson, 1982; Taatgen & Lee, 2003) or
chunking (Newell, 1990), the majority of these presuppose that any speedup in performance reflects a predominantly mental phenomenon, and as such either ignore or marginalize the dependence of the speedup on the specific activity in which it was observed. Contrary to this view, when Agre and Shrager examined the behavior of their subject video frame by video frame, they found that much of the observed improvement could be attributed not to a mental speedup, but rather to qualitative shifts in the interaction between the photocopier machine and the subject performing the activity. For example, after pressing the START button on the machine, there were three flashes of light (as the subject was making three copies of each page) lasting roughly 2.5 seconds. Following the third flash, the book could be moved and prepared to copy the next page. Over the course of the experiment the subject began to exploit this property of the machine, by walking to the right of the copier while she pressed the START button so as to be ready to collect the pages, and positioning her hands to manipulate the book before the third flash of the copier happens.

Numerous small changes such as this propagated throughout the task, altering the scheduling and timing of later operations, which altered the dynamics of the photocopier machine itself, which in turn further altered the subject’s pattern of interaction with the machine. The cyclical nature of these changes, which the authors termed *routine evolution*, account for the observed power law of practice, and yet cannot be explained by a mental speedup or other purely task-independent mechanism, as they critically depended on specific features of the photocopier machine and the human operators ability to interact with this device.

While the photocopier experiment of Agre and Shrager provides a compelling example of how embodied dynamics are exploited in naturalistic settings, understanding the mechanisms leading to the observed changes may require more controlled experimental paradigms. One such study is presented by Augustyn and Rosenbaum (2005). In this experiment, subjects were presented with two circular targets on a computer screen. The subjects were required to move the cursor to a starting position anywhere along a line connecting the two targets. After choosing a starting position, one of the two targets (randomly chosen) disappeared, and the task for the participant was to move the mouse cursor into the remaining target within a short response interval.
The key manipulation of this experiment was that the width of one of the targets was varied. By making one of the targets larger, it became easier and faster for the subjects to move the cursor into that target, based on a well-known relationship between target size and distance known as Fitts’ law (see MacKenzie, 1992). Given the goal of maximizing performance in terms of landing the cursor inside the target within the time period, optimal performance in the task requires accounting for Fitts’ law, as well as numerous low-level dynamics of the specific task (for example, subjects moved a small wooden disk on a table to control the movement of the cursor). In keeping with their predictions, Augustyn and Rosenbaum found that their participants demonstrated performance that was indistinguishable from the optimal predictions. They interpreted this finding in terms of a metacognitive control mechanism able to predict the sensory consequences of actions.

Gray and colleagues (Gray, Sims, Fu, & Schoelles, 2006) examined behavior in a more complex block copying task, where subjects had to replicate a pattern of eight colored blocks seen in one location of the computer screen at another location. A key feature of their task was that the original pattern of blocks (target window) and the workspace area where the pattern was reproduced (workspace window) were both covered by occluding boxes. To uncover the target or workspace window, the subject needed to move the mouse cursor inside the occluding box. As a between-subjects manipulation, the investigators added a lockout time to the target window that varied between 0 and 3.2 seconds, such that the subject had to move the mouse cursor inside the box and wait the lockout duration before the target pattern could be viewed.

As dependent measures, the investigators recorded the duration that subjects studied the target pattern, the number of blocks of the pattern that were successfully copied to the workspace following the first uncovering of the target window, as well as the total number of visits required to complete each trial. Subject performance according to each of these measures systematically varied between conditions, with the general finding that as the lockout time increased, subjects spent more time studying the target pattern on each visit, placed more blocks following that visit, and required fewer accesses of the pattern to complete each trial. These adaptations to the lockout period reflected a tradeoff between the ability to successfully study and recall all eight blocks on the one
hand, and the temporal and motor cost of re-accessing the pattern if some of the blocks are forgotten or not encoded.

In order to understand the adaptive nature of these findings, the investigators constructed an ‘ideal performer analysis’ in which the optimal memory encoding strategy was derived given the structure of the task (perceptual-motor costs) and quantitative predictions regarding human declarative memory performance (based on the ACT-R theory of declarative memory, J. R. Anderson et al., 2004). Far from being arbitrary, they found that the subjects’ compensation to the lockout period was essentially optimal given the capacities and limitations of the human memory system. Based on these and related experimental findings (Gray & Boehm-Davis, 2000), Gray proposed the soft constraints hypothesis, which argues that all else being equal, the guiding principle in the organization and coordination of interactive behavior is one of local least-effort, as measured by time. Although the authors did not offer a process-level theory of how this constraint resulted in their subjects’ specific patterns of memory encoding and rehearsal, they concluded that

"to some degree, humans have some metacognitive sense regarding how likely they are to remember something, given how much effort they are willing to spend memorizing it, and given the length of time they need to remember it. […] However, given the varied nature of demands on memory, it does not seem likely that this metacognitive tuning would yield an immediate, optimal solution to each new memory challenge. In the case of the Blocks World task, we found that participants required on the order of 10 trials to fine tune their strategies to match the demands of the experimental condition" (Gray et al., 2006, p. 471).

These findings and the authors’ conclusions provide further evidence for the exploitation of embodied task dynamics in routine interactive behavior. Note that, as with the photocopier experiment (Agre & Shrager, 1990), the dynamics being exploited are not purely external (the physical and temporal structure of the artifact) or purely internal (the dynamics of human memory activation and decay). The cognitive control mechanism adapts to and exploits both. From the perspective of achieving efficiency in behavior, it is not clear that the adaptive system should treat internal or external dynamics differently, mirroring the philosophical stance of active externalism (Clark, 2003).
Shin and Rosenbaum (2002) conducted an experiment in which a largely cognitive task, single-digit arithmetic, was combined with a perceptual-motor task in order to determine if and how their coordination would be achieved. In this experiment, subjects performed repeated mental arithmetic, for example adding $2 + 6 - 4 + 3$. However, in order to access the operands, the subjects were required to move the mouse cursor back and forth, alternating between small targets on the computer screen. This nested aiming-arithmetic task was designed to examine how participants would coordinate their cognitive activities with the perceptual-motor activities required to complete each trial. As a control, the investigators also had subjects perform an aiming-only task and an arithmetic-only task.

Although numerous researchers have examined the effects of multiple competing tasks on the ability to execute any one of them, the majority of these have employed dual-task experiments not as evidence for coordination, but rather as evidence for interference effects, such as the widely studied psychological refractory period (PRP) paradigm (for a review, see Pashler, 1994). In contrast to this, Shin and Rosenbaum assumed that there is some strategic control over the ordering and scheduling of operations, and that in the nested aiming-arithmetic paradigm subjects might exploit this capacity to improve efficiency. They hypothesized that subjects would perform arithmetic steps in parallel with moving the mouse to the next target. Further, if the time required to complete a calculation step was longer than the movement time, subjects should alter the properties of their movements so that they arrive at the next target (and receive the next operand) just as they finished the current calculation. This scheduling strategy would minimize ‘dead time’, as well as possible interference from having to store multiple operands in working memory.

In keeping with these predictions, it was found that subjects waited longer in the nested aiming-arithmetic task before moving the mouse to the next target, and the duration of the movements were also longer, compared to the aiming-only condition. The investigators phrased their results in terms of a strategic scheduling process in cognitive control. A stronger claim for their findings is that subjects had knowledge (either implicit or explicit) of their movement dynamics, the perceptual-motor costs
defined by the task, and the properties of their own mental arithmetic performance, and exploited this knowledge to improve their performance.

2.1.1 Summary of previous findings

This section has presented the results of several empirical studies that can be interpreted as evidence for adaptation to embodied task dynamics in routine interactive behavior. What other conclusions can be drawn from these findings? On the one hand, the results from these studies might not seem terribly profound—another way of stating their common finding is that human subjects performed well on simple and repetitive tasks. Given the vast capacities of human intelligence, for example the ability to communicate linguistically, the study of such simple tasks may not seem promising from the perspective of understanding human intelligence. On the other hand, it is crucial to recognize that the mechanisms by which this efficiency is achieved are not currently understood. Although humans are capable of learning from linguistic instruction, we are never taught how to interact with the world in terms of the intricate and low-level coordination demonstrated by the experiments presented in this section. Yet despite their lack of instruction, humans exploit intricately detailed knowledge of nearly everything around and within them: knowledge of the sequential and temporal structure of task artifacts, the limitations and capacities of the motor system, the ability to remember information for short periods of time, and the temporal properties of mental arithmetic.

Stating that humans adapt to and exploit embodied task dynamics is a parsimonious way of summarizing the results from the experiments presented here, but it does not explain what is meant by ‘adapt to’ or ‘exploit’, or how either can be realized in computational, let alone biological processes. Specifically, what is the mechanism that, through routine experience in the world, acquires knowledge of internal and external dynamics and produces as output efficient and coordinated behavior? Previous studies in the area have been largely silent on the issue of mechanism. Agre and Shrager (1990) referred to a process of ‘routine evolution’ but gave no details on its implementation or operation. Augustyn and Rosenbaum (2005) interpreted their results in terms of a largely unspecified metacognitive controller. Gray et al (2006) developed a reinforcement learning algorithm (using Q-learning) to derive optimal task performance, but noted it
did not represent a theory of human cognitive functioning, due in part to the unrealistically slow learning performance of the algorithm. Shin and Rosenbaum (2002) suggested only that cognition and perceptual-motor control are coordinated through a generic process of scheduling. As has previously been noted, “a sufficiently vague theory is consistent with any empirical data” (Gilbert & Shallice, 2002). In the next chapter a candidate mechanism intended to account for these findings will be proposed, based on learning and exploiting forward and inverse internal models of the embodied dynamics of a task. Preceding that, a brief review will be provided of the literature on internal models in sensorimotor control.

2.2 Research on internal models in sensorimotor control

The purpose of this section is to give a brief review of literature relating to a fundamental construct in motor control research known as an internal model. This literature is of direct relevance, as internal models will play a central role in all of the experiments and computational models that follow. An internal model “is a structure or process in the central nervous system that mimics the behavior of some other natural process” (Jordan, 1996). For example, an internal model of the arm might capture some systematic relationship between hand position in Cartesian space and forces applied to muscles in the arm. The nature of the relationship may depend on properties of the arm, such as its mass and moments of inertia, as well as external factors such as gravity and friction.

Internal models are typically further divided into forward and inverse models, depending on the causal direction of the relationship being modeled. A forward model takes as input the current state and control information of the modeled system, and produces as output the predicted state some point in the future. An inverse model, on the other hand, reverses the direction of this mapping, and instead takes as input some desired state and produces as output the control information necessary to achieve that state. Based on this relationship, forward and inverse models are useful for different purposes—forward models can predict or anticipate the consequences of actions, whereas inverse models are fundamentally concerned with producing the signals that will successfully control a system.
The existence of some form of inverse model in low-level motor and oculomotor control is rather well accepted (Mehta & Schaal, 2002), although many details regarding the neural implementation of these structures remain to be worked out (Green, Meng, & Angelaki, 2007). A commonly cited example of the capacity to exploit inverse models comes from the vestibulo-ocular reflex (VOR), which sends control signals to the eye muscles to compensate for movements of the head and body. This system is particularly well-studied due to the comparative simplicity of the neural mechanisms underlying empirically observed behavior. If the head begins to rotate by some angular velocity, the eyes need to counter-rotate with the same velocity in order to maintain a stable visual field. The mapping from desired angular velocity to oculomotor commands constitutes an inverse model of the eye-head dynamics. Since the timescale of this reflex is faster than the neural conduction time required for responding to visual or vestibular feedback, and indeed occurs even in complete darkness (du Lac, Raymond, Sejnowski, & Lisberger, 1995), the reflex must be carried out in an open loop manner, and according to most contemporary theories, using an internal inverse model.

In parallel to the growing evidence for inverse models, in recent years the evidence for forward models in motor control has also been steadily increasing. One difficulty in demonstrating the existence of a forward model lies in the fact that the output of such a structure, a prediction of future events, is not directly expressed in terms of overt behavior, but is rather used to shape or facilitate future behavior. Demonstrating the existence of forward models in brain circuitry is similarly difficult since such structures must be inferred from spike activity using an unknown coding scheme. In spite of this difficulty, a number of experiments have been conducted that provide indirect evidence for forward models. These experiments rely on the idea that the predictions from forward models can be used as surrogates in the face of periodic or absent sensory input (Mehta & Schaal, 2002; Wolpert, Ghahramani, & Jordan, 1995), anticipate changes in object (Blakemore, Goodbody, & Wolpert, 1998) or effector dynamics (Bhushan & Shadmehr, 1999; Conditt, Gandolfo, & Mussa-Ivaldi, 1997), or allow feedback-based control in the presence of slow or absent sensory input (Desmurget & Grafton, 2000; Flanagan, Vetter, Johansson, & Wolpert, 2003; Flanagan & Wing, 1997; Miall & Jackson, 2006). Other indirect support for the role of forward models comes from
computational models that have been successfully fit to bodies of experimental evidence (Coenen & Sejnowski, 1996; McKinstry, Edelman, & Krichmar, 2006; Wolpert & Kawato, 1998).

One particularly clear demonstration of the contribution of forward models in motor performance is a study conducted by Flanagan and Wing (1997). In their experiment, subjects grasped a manipulandum that was instrumented with force sensors, and the subjects were required to execute simple one-dimensional push-pull movements. The properties of the manipulandum were altered by imposing forces on the object, including inertial and viscous forces, as well as a combination of both. In order to prevent the object from slipping from grasp, it was necessary to impose additional grip force during changes in external forces imposed on the object. Grip force was measured for the subjects over the course of learning in the experiment. It was shown that initially, subjects reacted to changes in load force—their changes in grip force lagged behind changes in the acceleration force of the object. However, after a short amount of practice in the task, subjects’ changes in grip force preceded load force by 14ms.

Finally, the mathematical literature on stochastic optimal control theory provides a strong argument in favor of the use of forward models by the motor system. A more detailed and technical review of this literature as it relates to motor control is provided in (Todorov, 1998, 2007). In essence, this approach begins with the realization that human sensory information is fundamentally and inherently uncertain. Further, in the case of motor control, feedback on the outcomes of actions is delayed relative to their actual occurrence in the world, due to neural conduction delays. Results from control theory demonstrate that if a probabilistic model of the system state and dynamics can be acquired, then it is possible to not only optimally estimate the current state of the system, but also optimally control the system to achieve a desired consequence. The Kalman filter, which has been described as one of the most important mathematical results of the 20th century (Casti, 2000), represents the optimal solution to the combined estimation and control problem for the case of systems with linear dynamics and Gaussian noise. Recently, Saunders and Knill (2004) tested the extent to which human reaching movements could be explained by an optimal control model by using a movement task
in which hand position was perturbed at various points along the movement. The results of their experiment were consistent with a model based on a Kalman filter.

It is of particular interest that the implementation of a Kalman filter requires the availability of an internal model of the dynamics of the system being modeled. It will similarly be seen that for all of the optimality analyses conducted in this research, the computation of optimal performance requires the existence of the mechanism under investigation; that is, the existence of an internal model of the relevant embodied dynamics. This is not a matter of choice in how the optimal predictions are computed, but is rather a fundamental property of any optimal system that must act in the face of uncertain information. To the extent that human performance approaches optimality, we are given some measure of evidence that the mechanisms used by the cognitive system approximate the exploitation of an internal model.

2.3 Summary

This chapter has reviewed two literatures relevant to the hypotheses of this dissertation. First, it has reviewed recent studies conducted on the nature of human performance in routine interactive tasks. These studies have all shared the property of studying human performance when multiple processing systems (cognitive, perceptual, and/or motor) must all be coordinated towards the common achievement of some task or goal. While the results of these studies have documented the efficiency and intelligence of human behavior in routine interactive tasks, they have done little to elucidate the computational foundations for such performance. Secondly, this chapter has reviewed research on internal models in motor control. In this literature, internal models have played a number of behaviorally relevant roles, from sensory prediction to learning to manipulate novel tools or act in the face of externally imposed forces. It should be apparent from their separate treatment in this chapter that the two research agendas have been developed largely independently of each other. However, it will subsequently be seen that the concept of internal models can play an important role in developing theories of the unified cognitive system.
3. Learning and exploiting internal models of embodied dynamics

"A running program is the moment of truth" (Simon, 1992; p. 154)

Internal models are computational structures in the brain that capture and predict the dynamics of ongoing performance. From a computational perspective, internal models can serve numerous functional roles: forward models may be used for state estimation, anticipating the consequences of actions, filling in for missing sensory input, detecting errors in performance, ignoring self-generated sensory stimuli (sensory cancellation), and dealing with delayed feedback. Inverse models can be used to generate the behaviors necessary to achieve a desired state and compensate for predictable external disturbances.

It is without controversy that by the time we reach adulthood, humans must learn to perform an incredible variety of tasks. Most of these tasks are routine and performed on a daily basis (for example, opening doors, dialing telephones, etc.). Yet even novel tasks consist of a small and relatively stable set of low-level interactive routines: moving the hand to the point in space fixated by the visual system, encoding digits in working memory, etc. Rather than learning to perform each new task as an independent learning challenge, the rational response to the human task environment is to learn the embodied dynamics that are common across tasks. From an ecological perspective, acquiring an accurate internal model of one’s embodied dynamics while performing a task may be of greater utility than learning to perform the task itself.

Unfortunately, these dynamics are not entirely constant across tasks and across the individual’s lifetime. If that were the case there would be no need for learning, as evolution could provide accurate predictive models from birth. The properties of the motor system slowly change as humans develop—limbs grow, muscles strengthen and weaken, etc.—and the same is true of the cognitive and perceptual systems. However, these dynamics can also change on a short time scale. Working memory performance changes as a function of stress, sleep level, and caffeine intake, among other moderators (Gunzelmann, Gluck, Kershner, Van Dongen, & Dinges, 2007; Lieberman, Tharion, Shukitt-Hale, Speckman, & Tulley, 2002). Thus, internal models must be learnable in the lifetime of the individual, and must be adaptive to changes that occur over relatively short timescales.
This chapter first presents a review of recent computational theories of learning internal models in the motor domain. Reviews are also provided in (Jordan, 1996; Wolpert & Ghahramani, 2004). Several of the approaches are evaluated on a simple artificial data set (acquiring internal models of simple one-input, one-output systems). The construct of an internal model is then incorporated into a unified modeling approach called the adaptive internal model framework. Finally, a computational model of a simple motor planning task is developed using this framework, based on the idea of acquiring and exploiting internal models of the relationship between movement difficulty and movement time known as Fitts’ law.

3.1 Learning forward models

In formal terms, a forward model takes as input current context or state information $x$, and an action $a$, and generates as output the predicted consequences of that action, $y_{pred} \leftarrow f(x,a)$. For simplicity, each component of the system can be considered a single value, but the same approach is equally applicable when some or all of the components are vectors. The predicted output $y$ can represent sensory information or a prediction of the future state of the system. Learning a forward model proceeds by acting in the world and observing the resulting consequences, generating a sequence of $\langle x,a,y_{obs} \rangle$ exemplars. If the forward model represents a perfect model of the system, then one would expect that $y_{obs} = y_{pred}$. If the model were incorrect or incomplete, then one would expect there to be some discrepancy between the observed and predicted values. A common measure of this discrepancy is squared error,

$$ E = \frac{1}{2} (y_{obs} - y_{pred})^2 $$

$$ = \frac{1}{2} (y_{obs} - f(x,a))^2. $$

Different models or domains might employ different error functions that maximize criteria other than squared error. Learning proceeds by altering the forward model $f(x,a)$ in a manner that reduces this error. Learning algorithms that operate by reducing the discrepancy between an observed and a desired output fall into the class of algorithms known as supervised learning. A review of supervised learning in general is
beyond the scope of this thesis, but recent textbooks include (Mitchell, 1997) and (Bishop, 2006). Instead, this section will focus on one particular approach to supervised learning, namely feedforward neural networks. The focus here is on neural networks since they are universal nonparametric function approximators (Hecht-Nielsen, 1989), and are thus (in theory) capable of representing any desired function mapping inputs to outputs, and less importantly, because they more closely resemble the known structure of biological organisms than other approaches to supervised learning.

A typical neural network consists of three layers of processing units, referred to as input, hidden, and output layers. The layers are connected by two sets of adaptive weights. Each processing unit in a layer performs some fixed transformation of its inputs, \( \phi(\cdot) \). The outputs of each unit are linearly combined using the weights, and fed as input to the next layer. The overall operation of the network can thus be written as

\[
y = f(x,a,w) = \phi^{(3)} \left( \sum_{j=0}^{M} w_{kj} \phi^{(2)} \left( \sum_{i=0}^{N} w_{ij} \phi^{(1)}(x_i) \right) \right),
\]

where \( w_{kj} \) represents the weight parameter linking unit \( k \) and unit \( j \), and \( \phi^{(n)}(\cdot) \) is the output transformation for the units in layer \( n \). Figure 3-1 shows a graphical representation of the operation of a typical neural network\(^1\).

---

\(^1\) Omitted in this discussion, but included in all the simulations that follow, is the use of bias units, which are additional units in the input and hidden layers having a constant fixed output of 1. These units play an analogous role to the constant term in linear regression models.
Figure 3-1. Diagram of a typical feedforward neural network with two input units, three hidden units, and one output unit. Information flows from left to right through the network.

Training is performed by adjusting the weights in the neural network to reduce the output error, typically by computing the gradient of the error measure with respect to the network weights, $\frac{\partial E}{\partial w}$. The most widely used procedure for optimizing the weights in a neural network is known as error backpropagation (Rumelhart, Hinton, & Williams, 1986), which in its simplest form amounts to a steepest descent search. The term ‘neural network’ originates from attempts to explain how complex information processing may be carried out in biological systems using large numbers of simple, undifferentiated processing elements (e.g., neurons). However, soon after the publication of the earliest models, researchers argued that the backpropagation algorithm is biologically implausible (Grossberg, 1987; Stork, 1989). On top of its limited biological relevance, it turns out that the most commonly used form of error backpropagation (the steepest descent variant) is also a numerically poor algorithm (Bishop, 2006). More biologically plausible learning rules have been proposed (O'Reilly, 1996), but for the purpose of this research it suffices to assume that there exists some biological process by which the brain is capable of adjusting weights in a supervised learning fashion based on sample pairs of input and output data.

To demonstrate the capacity of neural networks to learn an arbitrary forward model mapping (hypothetical) actions and their outcomes, two artificial data sets were generated by sampling 100 points each from the functions $y = 4(x - 0.5)^2$ and
\[ y = x + 0.3\sin(2\pi x), \] and then adding uniform noise over the interval \((0.1, 0.1)\). Neural networks with one input unit, 4 hidden units, and a single output unit were trained on the data sets. A linear transformation function was used for the input and output layers, while the hidden layers used tanh activation functions. The neural networks were trained by minimizing a sum-of-squares error function using the BFGS\(^2\) nonlinear optimization algorithm (Nocedal & Wright, 1999). Each network was trained for 150 iterations, and no attempt was made to adjust the structure of the network (number of hidden units) to fit the data. Figure 3-2 (top row) shows the two data sets used, and the predictions from the trained forward models. Figure 3-2 (bottom row) shows the absolute prediction error for each of the two models, or the difference between the neural network output and the underlying function used to generate each data set. As can be seen from the graphs, the neural networks were able to learn accurate forward models of each function.

---

\(^2\) Broyden-Fletcher-Goldfarb-Shanno. The BFGS algorithm is an unconstrained optimization routine derived from Newton’s method, using an approximation of the Hessian matrix. The algorithm is numerically robust and typically performs much better than gradient descent, especially on difficult problems (Nocedal & Wright, 1999).
3.2 Learning inverse models

Forward models may be used for numerous practical and biologically relevant purposes, including state estimation, anticipating the consequences of actions, etc. Ultimately however, humans and adaptive agents are concerned with achieving desired states or outcomes in the world, and this requires the ability to generate the actions that will
achieve those states. While forward models can predict the consequences of actions, in isolation they cannot be used for behavioral control, as they do not provide the inverse mapping, or the mapping from desired states to the actions that will achieve those states. Internal models that capture this mapping are known as inverse models, for the reason that pairing a forward and an inverse model of the same system will yield an identity mapping, as shown in Figure 3-3.

![Figure 3-3. The reciprocal relationship between forward and inverse models. x is the current state information, a is an action, and y is a resulting or desired state.](image)

Acquiring inverse models through interaction with an external environment is a more difficult task than the one of acquiring a forward model. In the case of the forward model, the desired output of the system is the outcome that was actually observed. Thus the error signal used for training the model is readily available. In the case of an inverse model, the desired output is the optimal action for achieving a desired state. But if this optimal action were already available to the agent, there would obviously be no need for learning an inverse model. This section reviews three methods that can be used for learning or approximating an inverse model of a system.

### 3.2.1 Direct inverse modeling

One of the simplest approaches to learning an inverse model is to simply reverse the role of inputs and outputs used to train a forward model. That is, given an observation \( \langle x, a, y_{\text{obs}} \rangle \), treat the current state \( x \) and resulting outcome \( y_{\text{obs}} \) as inputs to the neural network, and train the network to reproduce the action \( a \) that yielded the observed outcome. Consider as an example an infant learning to reach for an object. Let \( x \) be the starting position of the hand, \( a \) be some quantity representing the motor actions executed, \( y_{\text{desired}} \) be the intended reaching target, and \( y_{\text{obs}} \) the actual outcome of the
infant’s motor commands. If on a particular attempt the infant misses by 6 inches to the left of the intended target, the infant hasn’t gained any information about how to reach the intended target \( y_{\text{desired}} \). But the infant did gain useful information about how to reach the location 6 inches to the left \( y_{\text{obs}} \). Thus learning proceeds in a non-goal directed manner (Jordan & Rumelhart, 1992), by learning how to achieve outcomes that were actually observed, rather than the outcomes that were intended or desired. This approach to learning, called direct inverse modeling, has indeed been used to model how a robot can learn the hand-eye coordination necessary to reach for objects in space (Kuperstein, 1988).

For linear systems, learning by direct inverse modeling will converge to the correct control policy for the system (Jordan, 1996; Wolpert & Ghahramani, 2004). However, for nonlinear systems a general problem emerges with this learning scheme. The problem relates to the fact that physical systems typically exhibit a many-to-one property, such that multiple actions, or even an infinite number of different actions, will produce the same system output. The challenge of learning an inverse model then becomes learning a one-to-many mapping. Direct inverse modeling attempts to learn the average action that yields a particular outcome. In general, there is no guarantee that the average of \( n \) optimal actions will be an optimal action, or even that it will yield a remotely reasonable policy.

Figure 3-4 demonstrates direct inverse modeling applied to the same artificial data sets used for learning a forward model. The roles of the input and output values have been reversed, and the networks are trained to learn the mapping from desired output to the input (control signal) that will achieve that state. The neural network architecture and training is identical to that used in the forward model problem. Learning by minimizing sum-of-squares error leads to a very poor control policy, as indicated by the bottom row of Figure 3-4.
Figure 3-4. Direct inverse modeling applied to two artificial data sets. (Top row) The input and output values have been reversed, such that the desired system output is plotted on the horizontal axis and the control signal is plotted on the vertical axis. In both cases, the model fails to learn an appropriate control policy. (Bottom row) Control error for the two inverse models. Control error is defined as the absolute difference between the actual and desired outcome resulting from following the learned control policy shown in the top row.

3.2.2 Distal supervised learning

Distal supervised learning (Jordan & Rumelhart, 1992) is a learning approach designed to overcome the one-to-many mapping problem. The approach is based upon the realization that forward and inverse models form an identity mapping when paired (Figure 3-3). The name distal supervised learning derives from the property that in many
natural tasks, the feedback available to the learning agent is provided in a distal coordinate frame that is not immediately useful for improving control. For example, the infant learning to reach is given feedback in terms of how far the hand was from the target, but not in terms of the motor torques and forces needed to correct this error. However, if the infant has already learned an approximate forward model, then this model can be used to bridge the gap between actions and their distal consequences. Thus the inverse and forward models are joined together and used in a composite learning system. If a forward model has already been learned, then an optimal inverse model is one that yields an identity mapping when paired with the forward model. An incomplete or incorrect inverse model will not yield an identity mapping, and this provides an error signal that can be used to adjust the inverse model while holding the parameters of the forward model fixed. This composite learning scheme is illustrated in Figure 3-5.

![Composite Learning System Diagram](image)

**Figure 3-5.** Using distal supervised learning to train an inverse model. The dashed line indicates the training signal used to adjust the inverse model.

Jordan and Rumelhart (1992) demonstrated that the composite system can learn an effective control policy even when the forward model is partially inaccurate or not fully learned. This arrangement also predicts that humans should acquire forward models before inverse models—that is, learn to predict a system before they are able to successfully control it. Indeed, some empirical evidence for this has been demonstrated in human motor control (Flanagan et al., 2003).

Performance of the distal supervised learning algorithm was evaluated using the same two artificial data sets used previously. For each data set, a forward model was first learned using a neural network and supervised learning (the forward model neural
networks are identical to those presented in Figure 3-2). Each neural network was then paired with an inverse model, which also took the form of a neural network with the same architecture. The parameters (weights) of the forward model were held fixed and the parameters of the inverse model were adjusted to minimize the sum-of-squared prediction error in the composite learning system, \( E = \frac{1}{2} (y_{\text{pred}} - y_{\text{obs}})^2 \). Figure 3-6 shows the performance of the resulting system.

For the quadratic function (left column), distal supervised learning was able to acquire an effective control policy, despite the one-to-many learning problem. It achieved this by settling on one of the two branches of the function, rather than attempting to learn the average control strategy across both\(^3\). However, for the more complex data set (Figure 3-6, right column), the algorithm failed to learn an effective control policy. Similar results were obtained as the number of training iterations varied from 150 to 5,000 and as the number of hidden units in the neural network was varied from 4 to 50.

\(^{3}\)On all simulations, the neural network settled on the upper rather than the lower branch.
Figure 3-6. Distal supervised learning applied to two artificial data sets. A forward model was first trained on each data set; this model was then used to train an inverse model. For the first data set (left column) the algorithm is able to acquire an accurate control policy. For the second data set (right column) the algorithm fails to learn an appropriate control policy, demonstrated by the large prediction error (bottom right graph).

To understand why this is the case, consider the learned control policy (Figure 3-6, upper right). To achieve a desired output of zero (on the x-axis), the learned control policy predicts that the system should output a value of approximately 0.7. This value corresponds to a local minimum in the original data set (Figure 3-2, upper right). Thus values slightly larger or smaller than 0.7 will lead to outputs that are further from the desired output of zero, and the training becomes stuck in this local minimum. This example demonstrates that distal supervised learning can fail to learn an effective control policy for systems containing local minima and maxima. Indeed, the inverse model
learned by distal supervised learning may even result in worse performance than an inverse model learned by direct inverse modeling.

### 3.2.3 Trajectory sampling

Perhaps the simplest approach for learning a control policy avoids learning an inverse model entirely. Instead, if the agent can learn an approximate forward model of the system, then it is possible to search (either iteratively or in parallel) the forward model for an action that produces the desired output. This scheme is similar to the technique known as trajectory sampling in dynamic programming (Sutton & Barto, 1998). In low-dimensional control spaces, it is possible to evaluate potential actions by sampling randomly from the entire control space and reporting the best action found (the examples in this section all involve a one dimensional control space). In higher dimensions, uniform random sampling quickly becomes impractical due to the curse of dimensionality—the probability of randomly choosing an optimal or near-optimal action decreases exponentially in the number of dimensions in the control space. In this case, heuristics, domain knowledge or other approaches such as instance-based retrieval may be used to generate a restricted ‘candidate neighborhood’ likely to contain an effective control signal. This restricted neighborhood can then be sampled either sequentially or in parallel to find the best control signal.

As a demonstration of the approach, trajectory sampling was applied to the same data sets used previously, and the results are shown in Figure 3-7. Forward models were first learned for each function using feedforward neural networks. For each desired system output, 100 candidate actions were generated uniformly from the interval (0,1) and evaluated using the forward model. The forward model prediction for each action was compared to the desired system output, and the action yielding the smallest error was used as the output of the ‘inverse model’. As shown in Figure 3-7, the trajectory sampling approach was able to generate an effective control policy for both functions. This is to be expected given the low prediction error in their respective forward models. In fact, arbitrarily low error could be obtained (assuming zero error in the forward model) simply by increasing the number of random actions sampled for each desired outcome. To avoid an impractical search involving the entire action space, statistics on
the output of the model could be maintained and then used as prior probabilities to bias the search to particular locations of the action space. The number of sampled actions could then be reduced over time as the agent gains experience in a particular task domain.

Figure 3-7. Control policies found by sampling from a forward model. For both functions, arbitrarily low error (relative to the error in the forward model) can be achieved by varying the number of actions sampled. In these examples, 100 random actions were sampled for each desired outcome.

3.3 Summary of learning algorithms

The ability to exploit internal models depends critically on the ability to learn them. The previous sections have defined two types of internal model, forward and inverse, and evaluated different algorithms for learning them on the basis of experience in the world. Forward models are used to predict the consequences of actions, while inverse models
are used to generate the actions that will achieve a desired outcome. The task of learning a forward model is much simpler than learning an inverse model, due to the many-to-one structure of most tasks. Section 3.1 demonstrated how neural networks could be used to learn forward models of two artificial data sets. Section 3.2 considered three different algorithms for learning inverse models of the same data sets. Each of the algorithms has previously been proposed as a model of human learning in the domain of motor control. It was shown that two of the algorithms (direct inverse modeling and distal supervised learning) exhibited serious limitations, even when evaluated using a simple artificial data set. The third approach (trajectory sampling) is a technique for approximating an inverse model without explicitly acquiring one. This approach was shown to perform the best on the artificial data sets evaluated, though its applicability to more complex learning problems is potentially limited due to the curse of dimensionality.

It should be clear from these results that the actual learning mechanisms used by the human brain remain largely unidentified, and indeed, outperform currently known algorithms. The focus of this research is not on the particular algorithm used to learn an internal model, but rather the role that internal models could play in the control of routine interactive behavior. In subsequent chapters, trajectory sampling will be used to approximate the functionality of an inverse model. This is simply because it was found to be the most robust approach for the domains under consideration. The exact learning algorithms used by the brain and their biological implementation is likely to remain a significant challenge for the foreseeable future. That is not to say that this research is unconcerned with human learning. There are two limiting factors that determine human learning: the limited experience from which a human must learn, and the amount of knowledge that can be extracted on the basis of this limited experience by a particular learning algorithm. In the face of an unknown learning algorithm, the approach adopted in this research is that of maximizing learning efficiency—that is, extracting as much information as possible from a fixed set of experience. This approach is taken with the recognition that it still pales in comparison to the flexibility and capacities of the human cognitive system.
3.4 Exploiting internal models for embodied cognition

In the previous section it was demonstrated how internal models could be learned and exploited to control simple one-input, one-output systems. As the primary objective was to demonstrate the relevant concepts and the operation of various learning algorithms, the chosen examples were deliberately abstract. In this section, it will be shown how forward models can be functionally combined in a unified framework for the control and optimization of behavior in routine interactive tasks. This framework will form the basis for the cognitive models implemented later in this chapter, and subsequent chapters of this thesis.

3.4.1 The adaptive internal model framework

In isolation, the construct of an internal model is insufficient to account for the human ability to achieve goals in the world. Trivially, an internal model must be paired with a sensorimotor system so that actions can be executed in the world and their consequences perceived. In addition, a memory system is useful for storing past interactions with the environment and retrieving relevant experience for the current task. The contents of memory form the basis for the internal model such that the model’s predictions are grounded in its previous experience in the world.

It must also be the case that actions are chosen based on their subjective value in the context of a given task environment. Consider a hypothetical adaptive agent that has acquired an internal model of the braking dynamics of an automobile—it ‘knows’ that slamming on the brakes will stop the car quickly (forward model), or that in order to stop the car quickly it must slam on the brakes (inverse model). But merely knowing the outcomes of actions does not indicate if and when those actions should be performed. This is simply because the utility of the same action can vary tremendously depending on the context of the current environment. In order for behavior to be adaptive for a given context, an agent must also be possessed of the ability to evaluate action consequences according to their subjective utility given the agent’s current goals.

Finally, it is necessary that some means is available for arbitrating between alternative actions. Given a set of potential actions and their expected value in the current setting, a naïve strategy is simply to select the action with the current highest
expected value. In natural environments this strategy is often suboptimal, as realistic agents invariably have limited knowledge about their environments. Therefore, the optimal action selection strategy is typically one that balances exploiting current knowledge of the environment with exploration of locally suboptimal alternatives in order to gather more information. In machine learning, this tradeoff is known as the exploration-exploitation dilemma (Sutton & Barto, 1998), and computing optimal action selection policies can be a deceptively difficult task even for trivial environments (e.g., Streeter & Smith, 2006). A simple heuristic that often works well is to sample actions probabilistically according to their currently estimated utilities, such that the action with the highest utility is most likely to be selected, while actions that have lower utilities have lower, but non-zero, probabilities of being selected.

Figure 3-8 illustrates a unified computational framework (henceforth referred to as the adaptive internal model framework) incorporating each of the elements just discussed. For a given state $s$, the action generation module produces a set of candidate actions $a$. These actions are conveyed to an internal model, as well as to the action selection module. The internal model generates predicted outcomes $\hat{s}$ for each action. These predicted outcomes are evaluated by the reward module, and the predicted reward $\hat{r}$ for each action is sent to the action selection module. The action selection module has two major roles, maintaining a value function, and performing action selection. The value function learns a mapping from states and actions to their expected values; this can be viewed as learning a model of the output of the reward module. Second, given a set of actions and a current state, it selects from the candidate actions based on the value function. This can be done by selecting the action with the highest expected value, using the soft-max function (Sutton & Barto, 1998), or some other approach. The selected action is then executed by the sensorimotor system, which perceives the resulting state of the environment. The memory system encodes current states and actions, as well as their resulting outcomes.
Figure 3-8. The adaptive internal model framework.

The dashed lines in the figure indicate training signals that can be used to adaptively tune the various components of the architecture for a given task. The memory system stores relevant experience for the current task; this experience can be used to adjust or train the internal model in a supervised learning fashion. The predicted rewards generated by the reward module can be used to train the value function based on reinforcement learning\(^4\). After executing an action, the difference between the predicted and observed reward \(\delta\) can also be used to adjust the action generation module: if an action yielded better than expected outcomes, the tendency to produce that action should

---

\(^4\) The implementation of the framework as used in this thesis does not incorporate learning a value function, as the primary theoretical focus is on the construct of learning and exploiting internal models.
be increased. Similarly, if an outcome is worse than expected, the tendency to generate that action should be decreased. This approach to adjusting an action control policy is equivalent to the actor-critic architecture in machine learning (Sutton & Barto, 1998).

3.4.2 Biological evidence for the adaptive internal model framework

Although the modeling framework just presented is not intended to be a biological theory of learning, the general structure of Figure 3-8 is broadly consistent with the organization, interconnections, and hypothesized functions of the cerebral cortex, the basal ganglia, and the cerebellum. To the extent that the computations performed by the human brain are understood, this knowledge should be used to help inform and constrain theories in cognitive science.

The cortical role in action generation has long been recognized, both in terms of motor commands generated by the primary motor cortex (Georgopoulos, 1991), as well as the role of the frontal cortex in the control of sequential behavior and higher cognition (Cooper, Schwartz, Yule, & Shallice, 2005; Humphreys & Samson, 2004). Cortical areas have also been implicated in the ability to evaluate the utility of the predicted consequences of actions (Bechara, Tranel, Damasio, & Damasio, 1996), corresponding to the hypothesized function of the reward module in Figure 3-8.

In recent years, numerous researchers have proposed that the basal ganglia is involved in the processing of reward information in the brain (Graybiel, 2005; Montague, Hyman, & Cohen, 2004; Schultz, 1998, 2006). In particular, the dopaminergic neurons originating in the substantia nigra of the basal ganglia respond to rewarding stimuli, regardless of the form of the reward (for example, food versus liquid reward given to animals). Larger rewards generally yield higher dopamine neuron activity, while negative reinforcement (pain or aversive stimuli) leads to suppression of activity in these cells. Thus, dopamine neurons projecting from the basal ganglia appear to encode the valence of reward, and not the content of the rewarding stimuli itself.

A complication of this story is that dopamine neurons do not encode reward value directly, but rather the prediction error for rewards. Better-than-expected rewards yield increases in dopamine neuron activity, while worse-than-expected rewards lead to suppression. Thus, dopamine neuron activity can be observed to decrease following the
delivery of a positive reward if the reward was smaller than anticipated, and no increase or decrease in dopamine activity is observed when a delivered reward was fully predicted. Finally, over the course of training it has been shown that the dopamine neuron response is gradually transferred back in time from the moment of reward delivery (when first learning a novel task), to the reward-predicting stimuli that precede future reward (Schultz, 1998). These results have been taken as evidence that dopamine activity represents the temporal difference (TD) error (Sutton & Barto, 1998) in the prediction of future reward.

Dopamine neurons from the basal ganglia project throughout the cortex, including prefrontal cortex (Strick, 2004). The temporal difference error encoded by these neurons (corresponding to the δ signal in Figure 3-8) can be used to train the action generation modules of the brain in a manner similar to the actor-critic architecture in reinforcement learning. The actor-critic approach has previously been proposed as a biological model of reinforcement learning in the brain (Houk, Adams, & Barto, 1995). In addition to the diffuse projections throughout the cortex, dopamine neurons also project recurrently to structures within the basal ganglia, particularly the striosome compartment of the striatum. It has been suggested that these projections enable the striosome to learn a value function that maps states and actions directly to their expected value, while the matrix component of the striatum is responsible for the action selection function of the basal ganglia (Graybiel & Saka, 2004). Inputs to the basal ganglia originate in areas throughout the cortex (Strick, 2004), however there are no direct connections between the cerebellum and basal ganglia. This fact motivates the structure in Figure 3-8 according to which the outputs of the internal model (in the cerebellum) are not fed directly to the action selection module, but rather are mediated through the reward module in the cortex.

An emerging consensus among researchers is that the cerebeller cortex and nuclei are the most likely candidates for the neural implementation of forward and inverse models. Although the cerebellum has long been thought to be responsible solely for generating motor control signals, with theories dating back to the 1700s (for a review, see Schmahmann, 2004), growing evidence strongly disputes this view. Although lesions or neurodegenerative disorders of the cerebellum do result in specific motor impairments
(Nowak, Timmann, & Hermsdörfer, 2007), these impairments are typically less severe than one would imagine if the cerebellum was the primary structure responsible for generating these motor commands. Indeed, the rather surprising fact that humans can recover relatively normal motor behavior following complete resection or removal of the cerebellum (Bower & Parsons, 2003) precludes its role as the primary brain site for action generation.

More recent theories of cerebellar function have expanded its role significantly. This is driven in part by gross anatomical characteristics, such as the fact that the cerebellum contains more neurons than all other brain areas combined (Ito, 1984), and detailed neuroanatomical studies showing that cerebellar afferents include all cortical regions (Courchesne & Allen, 1997), including prefrontal regions (Schmahmann & Pandya, 1995; Strick, 2004) associated with higher cognitive function. Alongside these anatomical findings, fMRI studies have also documented recruitment of the cerebellum in non-motor tasks such as decision making under uncertainty (Blackwood et al., 2004), working memory (Hayter, Langdon, & Ramnani, 2007), and time estimation (Xu, Liu, Ashe, & Bushara, 2006). Researchers have also documented numerous cognitive impairments that follow lesion or degeneration of the cerebellum, a pattern that has been collectively referred to as the dysmetria of thought (Gottwald, Wilde, Mihajilovic, & Mehdorn, 2004; Schmahmann, 2004). This combined evidence definitively rules out theories of the cerebellum as a strictly motor-control center, and points to the possibility that the cerebellum serves to facilitate processing throughout the brain. Such a role is consistent with the use of internal models in the proposed framework.

Adopting a different approach, Doya (1999) argued for an understanding of brain function that proceeds not by assigning particular roles to areas of the brain (for example, defining “sensory”, “motor”, and “cognitive” brain areas), but rather by examining the types of computations that each area is particularly well-suited to perform. Based on this analysis, it was argued that the cerebellum is uniquely specialized for supervised learning, while the basal ganglia is ideally suited for learning based on scalar reward signals.

In summary, the components of the adaptive internal model framework illustrated in Figure 3-8 were based on the functional roles they could serve in shaping and controlling
behavior, rather than their biological plausibility or necessity. Despite this fact, the organization of the framework is broadly consistent with the known structure and function of cortical areas, the basal ganglia, and the cerebellum. It will be shown in the next section that the framework is also able to account for skilled human performance in an interactive task, and provide a detailed computational account of the origins of this behavior whereas previously only vague verbal theories were offered.

3.4.3 Experimental data set

As a demonstration of the functionality of the internal model framework, this section will present a computational model of the simple reaching task studied by Augustyn and Rosenbaum (2005). In this paradigm, subjects were required to place a cursor anywhere along a line between two circular targets. After placing the cursor, one of the targets (chosen randomly) would disappear and subjects were required to move the cursor into the remaining target within a short time limit. The investigators varied the relative width of one of the target rings, in the range of 1:1 (equal width) to 4:1. This manipulation was intended to manipulate the difficulty of the reaching task. In particular, a common formulation of Fitts’ law (MacKenzie, 1992) states that the time to complete a reaching movement ($MT$) is related to the target distance $D$ and the width $W$ of the target according to

$$MT = a + b \log_2 \left( \frac{D}{W} + 1 \right),$$

where the logarithm term represents the index of difficulty of the movement. Movements of shorter distances, and to larger targets, can be completed more quickly than movements to smaller or more distant targets. Augustyn and Rosenbaum found that subjects in their experiment demonstrated remarkable sensitivity to the embodied dynamic of Fitts’ law, and adjusted the cursor start position to maintain an equal index of difficulty for both targets on the screen. This strategy minimizes the maximum reaching time (since the disappearing target was unpredictable), and is equivalent to achieving minimax optimal performance for the particular task. Figure 3-9 illustrates the paradigm (Figure 3-9a–b) and the main empirical results (Figure 3-9c). In Figure 3-9c, filled marker points indicate subject data, and the open markers are the minimax optimal
strategy computed based on Fitts’ Law. Circles correspond to trials on which the lower of the two targets was the smaller of the pair, while squares correspond to trials on which the lower target was the larger. Subjects in the experiment chose a cursor start position closer to the smaller (more difficult) target as the difference between the two target widths increased.

Figure 3-9. The reaching task studied by Augustyn and Rosenbaum (2005). (a) On each trial, the subject placed a cursor along a line separating two circular targets. (b) The line and one of the targets (randomly chosen) disappeared, and the subject was required move the cursor to the remaining target within a short interval. (c) Fractional distance of the chosen start position $D_L$ from the lower of the two targets, as a function of the targets’ relative widths. Error bars indicate $+/-$ 1 SE. Figure adapted from (Augustyn & Rosenbaum, 2005).

Based on the close correspondence between subject data and the optimal strategy computed based on Fitts’ law, the authors concluded that performance in the experiment was “consistent with a metacognitive account” (Augustyn & Rosenbaum, 2005, p. 915), but offered no insight as to the properties or capacities of such a metacognitive system. As an alternative hypothesis, the authors investigated whether subjects adapted to end-of-trial feedback. In their experiment, if subjects failed to move the cursor into the target within a short time interval, a buzzer was sounded. No evidence for trial-by-trial adaptation to this feedback was found in their data. Thus, the working definition of a
metacognitive control strategy offered by the authors is one that a) results in optimal performance and b) does not appear to be learned on the basis of feedback from the environment.

In this section, a more detailed account of subject performance is offered in terms of a computational cognitive model that learns to perform the task based on acquiring and exploiting internal models of embodied dynamics. The computational instantiation of the model offers a precise statement of the theory, and the ability for the model to match human performance proves the sufficiency of the theory to account for human performance.

3.4.4 Cognitive model based on the adaptive internal model framework

A cognitive model for the Augustyn and Rosenbaum (2005) experiment was developed based on the adaptive internal model framework illustrated in Figure 3-8. The cognitive model included an internal model that predicted motor movement durations based on target distance and width. This internal model took the form of a feedforward neural network with two input units, corresponding to target distance and width, and one output unit, corresponding to predicted movement time. The input and output units used linear transformation functions, while the hidden units used the tanh function. It is reasonable to assume that human subjects came in to the experiment with a lifetime of experience reaching for objects, and thus it is plausible that humans might already have an accurate internal model of the embodied dynamics of Fitts’ law at the outset of the task. However, for the purpose of the computational simulations, the agents started out the task with no background experience in motor movements or Fitts’ law.

On each trial of the experiment, a pair of targets was presented to the agent. The distance between the centers of the targets was fixed at 1.0 units\(^5\). The smaller of the two targets had a diameter of 0.1, and the diameter of the larger target was randomly drawn from the set \(\{0.1, 0.2, 0.3, 0.4\}\). For each trial, the action generation module simply generated 100 random actions, with an action corresponding to a point along a line

\(^5\) The unit of measure is arbitrary, as only the ratio of distances and widths matters in terms of optimal predictions.
separating the two targets, in the interval \((0,1)\). This action defined a distance to each of the two targets. The resulting distance and width for each target were sent to the internal model. The output of the internal model consisted of a predicted movement duration for each target. The reward module computed the maximum of the two predicted movement times, and returned the negative of this value as an expected cost. That is, the agent valued minimizing the maximum expected movement duration. The origin of this particular reward function is not modeled, as it is simply assumed that human subjects knew that a useful strategy was to minimize the maximum movement time based on the experimental instructions and their prior experience. The action selection module simply executed the action that had the highest expected value—\(i.e.,\) no noise or stochasticity in action selection was modeled.

After selecting an action, one of the two targets disappeared (randomly selected) and the sensorimotor system moved the simulated cursor to the remaining target. The movement time for this action was computed based on Fitts’ law. The movement time parameters \(a\) and \(b\) were arbitrarily set to 0 and 1, respectively, as their exact values do not influence the predictions or performance of the model (the optimal strategy is defined purely in terms of the index of difficulty term of Fitts’ law). The sensory system perceived the resulting movement distance, target width, and movement time. Each trial instance was stored in the memory system in the form of a \(<\text{target-width}, \text{target-distance}, \text{movement-time}>\) tuple, and no forgetting or memory decay was modeled. The contents of memory were used to re-train the internal model following each trial. In particular, half of the stored instances (randomly chosen) were used as a training set, and the remaining half were used as a validation set. The neural network was trained on the training set for 200 iterations using the BFGS (Nocedal & Wright, 1999) nonlinear optimization algorithm, and the network weights that resulted in the lowest prediction error on the validation set were used as the trained neural network weights. It is worth emphasizing that the neural network training procedure was not used to fit the model to human data, but rather to maximize objective performance in the task. No adjustments to the model structure or training procedure were made to fit its predictions to human data.

The use of a validation set could have been avoided by limiting the number of training iterations or adding a regularization term (Bishop, 2006) to the error function;
however, either approach would have introduced a free parameter into the model. Bayesian approaches to training neural networks have also been developed, however, the implementation of such approaches would have introduced significant computational demands, and offered nothing in terms of theoretical content of the model. By using a validation set, the learning performance of the model is somewhat impaired (as it only uses half of its available experience to learn from), but the predictions of the model are entirely parameter free.

Each agent completed 300 trials of the experiment, the same number as completed by the human subjects. In the case of the human experiment the first 12 trials were not analyzed, and so they are excluded from the following model analysis as well. Fifty model agents performed the task. Figure 3-10 compares human subject performance in the original experiment to the model performance according to the main dependent measure, the chosen starting distance to the lower of the two targets. As can be seen, the model data is nearly indistinguishable from the human data.

Figure 3-10. Comparison of human subject data (a) to a cognitive model (b) based on acquiring and exploiting an internal model of the embodied dynamics of Fitts’ law.

(Augustyn & Rosenbaum, 2005) only provided a graph of human performance, so no quantitative model fits can be given.
Learning performance for the model was assessed by computing the optimal starting location for each trial\(^7\), and then computing the mean square difference between each model’s selected action and the optimal action. Additionally, the first 12 trials were included in this analysis. Figure 3-11 shows the average learning curve for the 50 agents. Although improvements in performance extend beyond the first 12 trials, the agents reached asymptotic performance by around 25 trials.

![Figure 3-11](image)

**Figure 3-11.** Average learning curve for the cognitive model, defined in terms of the mean square difference between selected and optimal actions.

Learning curves were not reported for the empirical data. However, Augustyn and Rosenbaum did test whether mean square error significantly varied across blocks (a block consisted of 24 trials in their experiment), using a repeated-measures ANOVA.

\(^7\) For this particular paradigm, the optimal starting location is simply given by 

\[ D \left( \frac{W_1}{W_1 + W_2} \right) \], where D is the distance separating the two targets, and \( W_1 \) and \( W_2 \) are their widths.
with block number as a factor. In their analysis, no significant change in performance was found. Given the rapid learning demonstrated by the model at the beginning of the experiment, it is possible that humans also showed some measure of learning in the first block of the experiment that was undetected by their ANOVA. Based on their analysis of human performance, Augustyn and Rosenbaum argued that,

“participants appreciated the relations among movement distance, target width, and movement time conveyed by Fitts’ law and apparently used that knowledge to guide their choices of start positions. Evidently, the participants brought this knowledge into the lab from their prior experience, judging from the fact that their start position choices were optimal from the start of the experiment and did not depend on success or failure on the preceding trial” (p. 915).

Similar conclusions could be reached from analysis of the model data. In one sense, the model results can be interpreted as further evidence for the interpretation provided by Augustyn and Rosenbaum (2005). However, whereas they interpreted their findings in support of an unspecified metacognitive control construct, the modeling results of this section represent a quantitative, falsifiable prediction obtained from a narrowly defined mechanism.

### 3.5 Summary

The ability to exploit internal predictive models of embodied dynamics requires the ability to learn such models, as embodied dynamics can change over the lifetime, and even from moment to moment as humans change tasks and environments. This section has reviewed major computational approaches to learning both forward and inverse internal models. Generally speaking, acquiring a forward model is a much simpler task as the training data matches the causal, many-to-one structure of the environment, and an error signal is readily available. Learning inverse models turns out to be a much harder task computationally, as the environment does not directly specify an appropriate error signal. In addition, inverse models in general have to deal with the problem of learning one-to-many mappings. The two more successful approaches evaluated in this section, distal supervised learning and trajectory sampling, both relied on an approximate forward model to simplify the computational learning problem. In the case of trajectory
sampling, it was shown that it is possible to entirely do away with an explicit inverse model.

Other learning approaches have been discussed in the motor control literature; for example the MOSAIC model (Haruno, Wolpert, & Kawato, 2001) consists of multiple paired forward and inverse models. Employing this scheme enables the decomposition of highly nonlinear system dynamics into much simpler and smaller learning problems, each of which is tackled by a separate module. This modular decomposition approach can in fact be combined with any of the other learning schemes presented here. The modular decomposition of a complex function is similar to the general mixture-of-experts framework (Jacobs, Jordan, Nowlan, & Hinton, 1991; Jordan & Jacobs, 1994) in supervised learning. Another successful approach in motor learning theory is the feedback error learning scheme (for reviews, see Jordan, 1996; Wolpert & Ghahramani, 2004). Evidenced by the variety of frameworks that have been proposed, it is clear that even in the restricted domain of motor control the biological learning processes are still poorly understood (Miall & Jackson, 2006; Miall, Weir, Wolpert, & Stein, 1993). Thus, the focus of this chapter has been to argue for the utility of learning internal models as a functional construct and demonstrate the capacity of various learning schemes, rather than argue for the biological necessity of one particular learning algorithm over another.

The construct of an internal model was then embedded in the adaptive internal model framework. The necessity for this was the fact that merely knowing the outcomes of actions does not provide any guidance as to the value of those actions. The integrated framework contains an internal model and reward module, a sensorimotor system for acting in the world and perceiving consequences, as well as a memory system and an action selection/value function module. Although each of these modules was rather crudely specified, the general organization of the architecture was argued to be consistent with the known structure and function of the cerebellum, basal ganglia, and areas in the cortex. More importantly however, it was shown that the architecture was able to account for skilled human performance in a behaviorally relevant interactive task.
4. Bayes’ Ball experiment

Previous research has accumulated substantial evidence that the human brain acquires and exploits internal predictive models of the sensory and motor systems (evidence for this was reviewed in Chapter 2 of this thesis). These internal models aid in adapting and fine-tuning the performance of the behaving system towards the effective achievement of goals in the external world. However, relatively little work has been done to test whether this capacity extends to the more general challenge of predicting embodied dynamics—the performance characteristics and interactions among perception and motor control, but also cognition, while engaged in concert within the context of a specific task environment. The ability to predict cognitive dynamics would be of tremendous practical value to any agent that must act in the world, as ultimately all three elements (cognition, perception, and motor control) must be coordinated and deployed towards the achievement of goals in any naturalistic task.

The purpose of this experiment was to test whether humans can predict and subsequently exploit a particular cognitive dynamic, namely the estimation of short temporal intervals. Time estimation plays a critical role in nearly all aspects of everyday life. To take but one example, in driving an automobile, knowing how long one can safely take one’s eyes off the road requires the ability to accurately estimate how much time has elapsed since looking away, as well as knowledge of how quickly things might change given the current road and traffic conditions. Determining whether to stop for a yellow light similarly requires the ability to predict how long yellow lights typically last, and so on.

A particular feature of the human ability to estimate time intervals that has long been studied by psychologists is the finding that the standard deviation of an estimated interval increases linearly with the magnitude of the interval. This scaling property is known as the Weber law of interval timing (Gibbon, 1977; Staddon & Higa, 1999). Figure 4-1 illustrates the Weber law in human subjects performing an interval estimation task. The data is taken from an experiment conducted by (Ratikin et al., 1998). Subjects in this experiment were required to reproduce time intervals lasting 8, 12, or 21 seconds

---

8 A portion of this experiment was previously published as (Sims & Gray, 2008).
following training on a reference interval. The subjects read random digits out loud during the interval to eliminate explicit counting strategies. As the magnitude of the interval increased, the variability (standard deviation of the response distribution) increased linearly, in keeping with the Weber law.

![Figure 4-1](image-url)

**Figure 4-1.** Demonstration of the Weber law of interval timing. The three curves show the distribution of subject-estimated temporal intervals lasting for 8, 12, and 21 seconds. Note that the width of the distribution (its standard deviation) increases with increasing interval magnitude. The plot markers indicate human subject data, while the smooth curves are model predictions based on a psychological theory of time estimation. Figure taken from (Taatgen, Rijn, & Anderson, 2007).

Given the ubiquity of tasks and environments requiring the ability to estimate intervals, it seems plausible that humans would also be capable of representing or predicting the uncertainty inherent in their estimates. While previous research has largely focused on internal models of deterministic systems, for example the kinematical relationship between joint angles and hand position, for a physical agent acting in a physical world there is fundamental variability and uncertainty in any action performed. Rather than ignoring this variability or viewing it as a hindrance to optimal performance, the rational response to variability is to exploit it, that is, form internal predictive models that incorporate the statistics of variability. In many instances, including the case of interval timing, such variability is no less lawful than other aspects of human
performance. The goal of the experiment, then, was to develop a paradigm where, if humans were able to acquire such stochastic predictive models of their temporal dynamics, they would be able to exploit this knowledge to achieve more optimal behavior.

The Bayes’ Ball experiment consisted of a small ball displayed on a computer screen. After pressing down a key on a response pad, the ball began moving horizontally from left to right with constant velocity. After traveling a short distance, the ball moved behind an occluding window that prevented the subject from viewing its position. When the key was released, the ball immediately stopped moving, and its position was revealed to the subject. Figure 4-2 shows the experimental apparatus used for the task. The paradigm was inspired by similar experiments (Maloney, Trommershäuser, & Landy, 2007; Trommershäuser, Landy, & Maloney, 2006; Trommershäuser, Maloney, & Landy, 2003) designed to assess implicit knowledge of the uncertainty inherent in reaching movements. In this version of the task, the theoretical focus is not on internal models of motor dynamics, but rather the cognitive dynamics of temporal interval estimation. A game-like interface was used to make it less obvious to subjects that the task was on interval estimation, and the rapid nature of the task was designed to mitigate the potential contribution of explicit counting strategies.

The experiment consisted of two phases, a training phase and a test phase. During the training phase, a small target was displayed on top of the occluding window, and the task for the participant was to time their response (key release) so that the hidden ball stopped exactly at the center of the target. The purpose of the training phase was to allow the participants the opportunity to acquire a predictive model of the combined temporal estimation and ball movement dynamics. No rewards or penalties were given during this phase of the experiment. During the test phase, the target was replaced by a configuration of three point regions (Figure 4-2b-d). Stopping the ball inside different regions yielded differing amounts of points for the participant (either +100, 0, or -200 points). The goal during the test phase was to accumulate as many points as possible. On each trial, the center region was worth +100 points. On some trials the two surrounding regions were penalty regions (each yielding -200 points), while on other trials only one
of the surrounding regions yielded a penalty. In addition, targets were presented at
different distances, requiring subjects to estimate shorter and longer intervals.

As a consequence of the asymmetrical penalty configuration, optimal performance
in this experiment required that subjects aim not for the center of the reward region, but
rather shifted slightly in the direction away from the penalty region. The magnitude of
the optimal shift depends on both internal and external dynamics. The internal dynamics
consist of the Weber law of interval timing, while the external dynamics include the
properties of the ball (its velocity), the target distance, and the task-specific point values
associated with stopping the ball in different regions.

4.1 Hypotheses

Given the prevalence of tasks requiring the ability to estimate short time intervals, it was
hypothesized that subjects would in fact be able to acquire and exploit an internal
predictive model of the dynamics of the Bayes’ Ball task. In terms of behavioral
measures, subjects were expected to compensate away from the center of the targets on
trials with asymmetrical penalty configurations. Further, it was hypothesized that the
magnitude of compensation would be greater for far targets than for near targets.
Violating either of these hypotheses would reveal a failure on the part of the subjects to
fully anticipate or adapt to the dynamics of the Weber law. If subjects had no awareness
of uncertainty in their time estimates, they would be expected to aim for the center of the
target region regardless of the surrounding penalty regions. If subjects possessed some
limited knowledge of their uncertainty, but not the linear scaling relationship of the
Weber law, they would be expected to compensate away from penalty regions, but not
increase the magnitude of their compensation as the target distance increased.

A stronger hypothesis was that the magnitude of compensation would be optimal
given the level of uncertainty specific to each subject. This hypothesis was based on a
series of previous studies demonstrating that humans were able to optimally compensate
for motor noise in a simple pointing experiment (Maloney et al., 2007; Trommershäuser
et al., 2006; Trommershäuser et al., 2003). To evaluate the possibility that human
subjects would achieve optimal performance in the Bayes’ Ball paradigm, an ideal
performer analysis was conducted for each subject in order to compare actual human performance to the computed optimal predictions.

Finally, it was hypothesized that human performance would reflect the acquisition of an internal model of the embodied dynamics of the task, and could not be explained by an alternative account of skill acquisition, namely reinforcement learning. To test this hypothesis, two cognitive models were developed that learned to perform the task: one that acquired and exploited an internal model, and the other based on learning the reward structure of the task (e.g., reinforcement learning; Sutton & Barto, 1998), but not an internal model of the embodied dynamics. These cognitive models were used to test the computational sufficiency of the internal model hypothesis, and reject reinforcement learning as a viable alternative explanation.

4.2 Method

4.2.1 Participants

Twenty-three undergraduate students at Rensselaer Polytechnic Institute volunteered to participate in the experiment in exchange for course credit. The subjects’ mean age was 19.4 years (SD = 1.2). Eight of the subjects were female. All subjects reported normal or corrected-to-normal vision.

4.2.2 Apparatus and materials

The experiment was conducted on an Apple G4 computer running version 10.4 of the Macintosh operating system. All experimental stimuli were displayed on a 17” LCD monitor set to 1280×1024 screen resolution. Participants interacted with the experimental software by pressing a key on a hardware response pad (Cedrus® model RB-834). The response pad contained circuitry for measuring and recording response times with 1 millisecond temporal resolution9. The experimental software was implemented in the lisp programming language using LispWorks® version 4.4.

9 In contrast, USB input devices such as keyboards and optical mice typically have a sampling rate of 125 Hz, or a temporal resolution of 8 ms (Cernich, Brennana, Barker, & Bleiberg, 2007).
On each trial of the experiment, a small blue ball (diameter = 27 pixels, 7.1 mm) was displayed near the left side of the screen. After pressing and holding down a key on the response pad, the ball began moving horizontally from left to right with a constant velocity of 1.0 pixels/ms. Trial-to-trial variation in the ball’s velocity was measured post-hoc and found to be negligible (SD = 0.004 pixels/ms). After traveling for approximately 200 ms, the ball reached a gray rectangle that acted as an occluding window. At this point, the ball continued to travel at the same velocity, but was hidden from the subjects’ view. Henceforth, after 200 ms subjects received no visual or other feedback as to the ball’s location until the end of the trial. The ball continued to move at constant velocity until either the subject released the response key or the ball reached the right side of the screen. At this point, the ball stopped and its terminal location was revealed to the subject for 500 ms.

The experiment consisted of two types of trials: training trials and test trials. During the training trials, a small red ‘x’ was displayed as a target at one of three fixed distances (distance = 77.71, 127.76, or 211.53 mm) from the start of the occluding window, as shown in Figure 4-2a. The task for the subject was to time their response (key release) such that the ball stopped at the center of the red ‘x’. After releasing the response key, subjects were shown the ball’s stopping location as feedback.
Figure 4-2. Apparatus used in the Bayes’ Ball task. (a) During the practice trials subjects attempted to stop the ball directly on the red ‘x’. (b) During the test trials, subjects attempted to maximize their point score by stopping in the reward region (green) and avoiding the penalty region(s) (red). This example shows the penalty-after configuration. (c) The penalty-before condition, and (d) the both-penalty condition.

During the test trials, subjects attempted to maximize a point score by stopping the ball in one of three color-coded regions on the screen. The center region was always worth 100 points and was represented by a green rectangle. On both-penalty trials, the regions to the left and right of the center region yielded a loss of 200 points (Figure 4-2d.) On penalty-before trials, the region to the left of the center region yielded a loss of 200 points while the region to the right was worth 0 points (Figure 4-2c). This mapping was reversed for penalty-after trials (Figure 4-2b). In addition, if the ball reached the right side of the screen it was counted as a penalty. The width of the center region was 21.86 mm, and the distance from the ball start position to the middle of the center region varied across trials (distance = 77.71, 127.76, or 211.53 mm). After each test trial, subjects received visual feedback regarding the ball’s stopping position, the point outcome for the most recent trial, as well as their total accumulated score. This feedback
was displayed for 500 ms. Auditory feedback was also given; on trials with a negative point outcome, a brief WAV recording of a buzzer was played through built-in speakers in the monitor, while a cheerful beep was played following trials resulting in a positive score.

### 4.2.3 Design

There were no between-subject conditions in the experiment. Subjects first completed 225 practice trials, grouped into blocks of three trials. Each block contained one trial at each target distance (*near*, *middle*, and *far*), presented in random order. Following the practice trials, subjects completed 450 test trials. Test trials were grouped into blocks of nine, with each block containing one trial at each combination of target distance and penalty configuration, presented in random order. Following the test trials, subjects were told to take a short break. The experimental software required that the experimenter enter a password before the subject could continue with the experiment. The purpose of the password was to ensure that all subjects took a short break at the same point during the experiment. Following the break, subjects completed another 75 practice trials, followed by a second set of 450 test trials. At the end of the experiment, subjects completed a short debriefing questionnaire (a sample questionnaire is given in Appendix A). The experimental design is summarized in Figure 4-3.

![Figure 4-3. Summary of experimental design used for Bayes' Ball experiment 1.](image-url)
4.2.4 Procedure
Subjects were tested individually in quiet rooms. Each subject was given a brief (approximately 10 minute) slide presentation that explained the task and the operation of the response pad. Subjects were informed that the ball moved at the same speed on every trial. For the practice trials, subjects were instructed to stop the ball as close as possible to the red ‘x’. For the test trials, subjects were instructed with the point value of each region, and were told to maximize their total point score. No strategy or advice was given to the subjects regarding how they could maximize their score. Subjects were informed that their scores would be recorded and added to a high score list at the end of the experiment. This was done to further incentivize subjects to perform to the best of their abilities.

Subjects began the experiment following completion of the slide presentation. Trials were self-paced, and given the velocity of the ball, each trial lasted a maximum of 1200 ms. The entire experiment was conducted in a single session and lasted for approximately 45 minutes.

4.3 Results
Three sets of results are presented in this section: results from the practice trials, results from the test trials, and results based on the post-experiment questionnaire. For the practice and test trials, subject data is excluded for all trials on which the ball reached the right edge of the screen. On instances when this happened, the experimental software terminated the trial without waiting for the subject to release the response key; consequently these trials do not provide a valid measure of subjects’ response time. Such trials accounted for 0.68% of the practice trial data and 0.49% of the test trial data.

4.3.1 Performance during practice trials of the experiment
During the practice trials of the experiment, subjects were instructed to aim for the center of the red ‘x’. The location of the ‘x’ varied according to the three target distances from trial to trial, subject to the constraint that the target appeared in each location once in each block of three trials. The three target distances are defined as near, middle, and far. Subject performance was quantified by defining the measure delta as the difference
between the ball’s terminal position on a given trial and the center of the target on that trial,

\[ \Delta = x_{\text{ball}} - x_{\text{tgt}}. \]

According to this measure, positive values of delta indicate stopping to the right of the target, and negative values indicate stopping to the left of the target. The subjects’ mean delta for the near, middle, and far targets was 0.08, -6.01, and -21.32 mm, respectively. A repeated-measures analysis of variance (ANOVA) of mean delta by target distance yielded a significant effect of distance, F(2,44) = 45.6, MSE = 40296.36, p < .001. This effect was unexpected, as subjects were ostensibly attempting to achieve a delta value equal to zero for each target distance. Instead, subjects increasingly stopped to the left of the target as the distance to the target increased. Inspection of data from individual subjects revealed that this bias was common to all subjects to varying degree, rather than being limited to a few isolated subjects.

To assess whether subject performance in the Bayes’ ball experiment conformed to the Weber law of interval timing, the standard deviation in delta was computed separately for each subject at each target distance. Weber’s law would predict that standard deviation should increase linearly with increasing target distance. A repeated-measures ANOVA of the subjects’ standard deviation by target distance yielded a significant effect of distance F(2,44) = 37.61, MSE = 35021.70, p < .001. Standard deviation in subjects’ targeting performance indeed increased with increasing target distance. As shown in Figure 4-4, this increase in standard deviation followed an almost perfect linear relationship with target distance, as would be expected if subjects’ performance were constrained by Weber’s Law.
Finally, recall that the practice trials were split into two phases, with the bulk of practice trials occurring at the beginning of the experiment, and 75 practice trials occurring immediately after the break at the midway point of the experiment. In order to assess whether performance changed over this span, the final 75 practice trials of phase 1 were compared with all 75 practice trials of phase 2. A repeated-measures ANOVA of standard deviation by practice phase (1 or 2) and distance (Near, Middle, or Far) yielded no significant main effects or interactions (all $p > 0.05$). Thus, subjects did not appear to significantly improve their targeting performance (decrease their standard deviation) across the two practice phases of the experiment. This suggests that subjects had reached asymptotic performance by the end of the first practice trial phase of the experiment, and before beginning the test phase.

4.3.2 Performance during test trials of the experiment

4.3.2.1 Mean score

In the previous section, it was shown that subjects exhibited a pattern of variability in their ability to estimate the ball’s position consistent with the Weber law of timing. During the test trials of the experiment, each trial resulted in a numerical point outcome (either $-200$, 0, or $+100$ points). If subjects were able to form a predictive estimate or internal model of the relationship between temporal intervals and the variability in the
ball’s resulting position, they should have been able to exploit this knowledge in order to improve their resulting score. If, on the other hand, subjects had no knowledge of any variability in their ability to estimate time intervals, it might be expected that subjects would aim for the center of the reward region regardless of the surrounding penalty configuration. Another possibility is that subjects might have some awareness of uncertainty in their temporal estimates, but lack detailed knowledge of the linear scaling of their uncertainty with increasing interval magnitude. In this case, subjects would aim for a point shifted away from the center of the reward region, but the amount of their compensation would not increase for the farther targets.

On average, subjects earned 12.68 points per trial (SD = 21.56), out of a theoretical maximum of 100 points per trial, suggesting that the task was rather challenging for the participants. Indeed, 6 out of the 23 subjects ended the task with negative scores. The overall mean scores for each target distance and penalty configuration are provided in Table 4-1. Subjects scored the lowest on both-penalty trials compared to penalty-before and penalty-after trials. This is to be expected, as on both-penalty trials there are two regions yielding negative scores, while on other trials there is only one penalty region. Also to be expected, subjects had more difficulty with the far targets than the near targets, with performance on the middle targets falling between the two.

In order to assess the extent of learning over the course of the test phase of the experiment, trials were aggregated into four epochs of 225 trials each. The mean scores for each epoch were compared using a one-way repeated measures ANOVA with epoch as factor. The results of this analysis indicated a significant change in score over the experiment (F(3,66) = 13.179, MSE = 865.90, p < .001). In particular, the mean score in epochs 3 and 4 were significantly higher than the score in epoch 1. The score in epoch 4 was not higher than the score in epoch 3, suggesting that any improvements in performance were confined to the first half of the experiment. The mean score for each epoch of test trials is plotted in Figure 4-5.
Figure 4-5. Mean test trial score across epochs of the experiment. Each epoch consists of 225 trials. Error bars indicated 95% confidence intervals.

4.3.2.2 Delta

For the test trials, delta was defined as the difference between the ball’s position and the center of the green reward region (similar to the error measure defined for the practice trials). By this measure, positive values indicate stopping to the right of the center of the reward region, and negative values to the left. The mean delta value for each distance and configuration is reported in Table 4-1 and plotted in Figure 4-6. A repeated-measures ANOVA comparing delta using target distance and penalty configuration as factors yielded a significant interaction between configuration and distance, $F(4, 88) = 20.713$, $\text{MSE} = 1933.06$, $p < 0.001$. This interaction was expected, as there should be little or no change in relative stopping position for the both-penalty configuration across the three target distances, whereas delta should increase with increasing target distance for the penalty-before trials, and decrease with increasing distance for the penalty-after trials. Unsurprisingly, the main effect of penalty configuration was also significant, $F(2, 44) = 61.365$, $\text{MSE} = 21620.81$, $p < .001$. Subjects shifted to the left (negative delta) for penalty-after trials, and shifted to the right (positive delta) for penalty-before trials.
The main effect of distance was also found to be significant, $F(2,44) = 39.16$, MSE $= 15155.15$, $p < .001$. This finding was unexpected, as on average, delta values from penalty-before and penalty-after trials should cancel at each distance. Across all three penalty configurations however, the relative stopping position shifted from positive for near targets, to negative values for the distant targets. Subjects thus appeared to overshoot the center of the target on the near trials, and undershoot the target on the far trials, regardless of penalty configuration. A similar bias was noted for the practice trials, where subjects consistently undershot the far targets.

Table 4-1. Mean target-relative stopping position (delta) and score as a function of target distance and penalty configuration.

<table>
<thead>
<tr>
<th>Distance</th>
<th>Config.</th>
<th>Mean score</th>
<th>+/- 95%CI</th>
<th>Delta (mm)</th>
<th>+/- 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Near</td>
<td>Penalty-before</td>
<td>61.30</td>
<td>5.94</td>
<td>6.1</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>Both-penalty</td>
<td>32.44</td>
<td>14.49</td>
<td>3.01</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Penalty-after</td>
<td>54.26</td>
<td>10.49</td>
<td>0.62</td>
<td>0.99</td>
</tr>
<tr>
<td>Middle</td>
<td>Penalty-before</td>
<td>21.52</td>
<td>9.80</td>
<td>4.35</td>
<td>2.08</td>
</tr>
<tr>
<td></td>
<td>Both-penalty</td>
<td>-19.22</td>
<td>16.42</td>
<td>-0.76</td>
<td>1.49</td>
</tr>
<tr>
<td></td>
<td>Penalty-after</td>
<td>28.74</td>
<td>9.54</td>
<td>-3.98</td>
<td>1.53</td>
</tr>
<tr>
<td>Far</td>
<td>Penalty-before</td>
<td>-2.78</td>
<td>11.56</td>
<td>1.79</td>
<td>2.03</td>
</tr>
<tr>
<td></td>
<td>Both-penalty</td>
<td>-72.44</td>
<td>14.79</td>
<td>-3.07</td>
<td>1.94</td>
</tr>
<tr>
<td></td>
<td>Penalty-after</td>
<td>10.30</td>
<td>7.18</td>
<td>-12.36</td>
<td>2.66</td>
</tr>
</tbody>
</table>
4.3.2.3 Compensation

To untangle this bias in stopping position from the main question of interest—compensation in stopping position due to uncertainty—an additional measure of subject performance was defined by subtracting each subject’s mean stopping position on both-penalty trials from the ball position on penalty-before and penalty-after trials. By subtracting the mean stopping position on both-penalty trials, this measure removes the bias in stopping position that is unrelated to penalty configuration. In addition, the sign of the difference was reversed for penalty-after trials, so that positive values of this measure consistently indicate shifting away from the penalty region, while negative values indicate shifting towards the penalty region, relative to stopping position on both-penalty trials. This measure, referred to as compensation is formally defined as

$$compensation = \begin{cases} 
(x_{ball} - \bar{x}_{both-penalty}) & x_{ball} \in \text{penalty-before} \\
-(x_{ball} - \bar{x}_{both-penalty}) & x_{ball} \in \text{penalty-after}
\end{cases}$$

Mean compensation is plotted in Figure 4-7. A repeated-measures ANOVA on compensation by target distance and penalty configuration yielded a significant interaction of distance and configuration, $F(2,44) = 8.714$, $MSE = 1871.08$, $p = .001$. Post-hoc analysis revealed that this interaction was due to a significant difference in compensation between penalty-before and penalty-after trials at the farther target distances. The main effect of distance was also found to be significant, $F(2,44) = 28.174$, $MSE = 3242.43$, $p < .001$. Thus, subjects increased their compensation for farther target distances, as would be expected if the subjects acquired internal predictive models that reflected the Weber law of timing.
Figure 4-7. Compensation relative to mean stopping position on both-penalty trials for each target distance. Error bars indicate 95% confidence intervals.

A further question of interest is the extent to which individual differences in compensation can be accounted for by individual variability in the ability to estimate time intervals. For example, subjects with high accuracy would be expected to show smaller compensation than subjects with low accuracy. Addressing this question is also of interest, as it is theoretically possible that the effects demonstrated so far have nothing to do with subjects’ variability, but rather the perceptual properties of the task environment (for example, farther target distances may cause subjects to increase compensation in some manner unrelated to temporal estimation variability). To address this question, a linear regression was performed using the subjects’ standard deviation on both-penalty trials as a predictor variable and mean compensation (the average of compensation on penalty-before and penalty-after trials) as the dependent measure. The regression was performed separately for each target distance, therefore eliminating target distance as a potential explanation for changes in compensation. For the near targets, individual differences in standard deviation were positively correlated with the amount of compensation, $r^2 = 0.23$, $t = 2.526$, $p < .05$. That is, subjects with higher variability in their targeting performance compensated further away from the penalty regions. However, for the middle and far targets, the correlation did not reach significance, both $p > .05$. Thus, there is at least limited evidence that subjects’ compensation in the task was related specifically to their individual ability to estimate temporal intervals, and not due to perceptual features or other extraneous properties of the task.
4.3.3 Post-experiment questionnaire

Following the experiment, each subject completed a short questionnaire with the questions read and answers recorded by the experimenter. The entire questionnaire is given in Appendix A. The analysis here focuses on questions 4 and 5. Question 4 asked if subjects consistently used any type of verbal strategy, such as counting out loud or counting silently to estimate the ball’s position. This question is relevant as the Weber law is based on implicit time estimation, and the same relationship does not necessarily hold if subjects adopt an explicit counting strategy. Question 5 asked subjects if they moved their eyes across the screen to simulate or track the ball’s position. This question is similarly relevant, as adopting such a strategy may (potentially) alter a subject’s performance.

Out of twenty-three subjects, five reported using a verbal strategy, while fourteen reported consistently using eye movements to attempt to track the ball’s location. Although subjects reported using these strategies, it is possible that their introspective accounts did not accurately reflect their actual strategy during the task, or that interval estimation consistent with the Weber law dominated their performance despite their attempts to adopt an explicit strategy. To address this issue, regression analyses were conducted with eye movement and verbal strategy questionnaire responses as predictor variables. Separate regression analyses were conducted to investigate whether these predictors influenced mean trial score or standard deviation in stopping position on both-penalty trials for each of the three target distances. The results of the analyses indicated no evidence that either a reported eye movement or verbal strategy influenced either subjects’ scores or standard deviations in resulting ball position (all \( p > .05 \)). Thus, despite different subjects reporting that they used different strategies, there is no available evidence to suggest that their strategies differentially influenced how they performed in the task.

4.3.4 Summary and discussion of results

The primary purpose of this experiment was to test the hypothesis that subjects would be able to acquire and exploit internal predictive models of their uncertainty in estimating short time intervals. According to the Weber law of timing, the ability to estimate such
intervals is characterized by a specific pattern of variability, with standard deviation in accuracy increasing linearly with the magnitude of the interval to be estimated. In the Bayes’ ball environment, estimating the ball’s position requires estimating how much time has elapsed since the ball disappeared behind the occluding window. As demonstrated by subject performance in the practice trials, uncertainty did increase with increasing target distance, and the relationship between the two was almost perfectly linear.

The test trials of the experiment introduced gains and losses associated with different regions surrounding the target. The asymmetrical penalty configurations (penalty-before and penalty-after) introduced the possibility of behaviorally measuring the extent to which subjects are aware of the uncertainty inherent in their ability to estimate time intervals. Empirically, it was found that subjects significantly deviated away from the center of the reward region, and in the direction opposite to the penalty region. Further, the amount of compensation significantly increased with increasing target distance. The best explanation for this pattern of results is that subjects were able to acquire and exploit an internal predictive model of the particular embodied dynamic relevant to performance in the Bayes’ ball task, namely the Weber law relation between temporal interval magnitude and estimation uncertainty.

Beyond these findings, several questions remain. First, to what extent is subject performance optimal in the Bayes’ ball task? Answering this question is of practical importance for several reasons. First, an increasing number of theories in cognitive science are explicitly based on the notion that human cognition can be productively understood as an optimally adapted system given a set of known constraints (e.g., on perception, motor control, memory, or other well-characterized constraints) and the statistics of the tasks and environments for which human intelligence evolved. For example, the ideal observer analysis approach (Geisler, 1989, 2003) in psychophysics is based on the assumption that the human visual system exploits all available information in an optimal manner. Models based on this approach have been used to explain human motion perception, and can account for apparent visual ‘illusions’ that occur under artificial or restricted viewing conditions (Weiss, Simoncelli, & Adelson, 2002). Anderson’s rational analysis framework (J. R. Anderson, 1990) has similarly used this
approach to show that the human memory system, and specifically the rate of forgetting, is nearly optimal given the statistical demands imposed by natural environments. Gray and colleagues (Gray et al., 2006) developed the *soft constraints hypothesis* based on the idea that the cognitive control system is optimized according to a principle of least effort in selecting low-level interactive routines. Each of these theoretical frameworks is based around the notion of optimality in everyday behavior.

With respect to the Bayes’ ball task environment, it is certainly the case that human subjects came in to the experiment with a lifetime of experience in observing and estimating short time intervals. It is also likely that there are situations common to the natural environment where uncertainty in time estimation can have serious consequences for the individual. Given that the Bayes’ Ball paradigm mirrors this aspect of the natural environment, it is a plausible hypothesis that human performance in Bayes’ ball might approximate optimal performance. If it is the case that humans are found to be suboptimal, the nature of the deviation from optimality may be used to gain further insight into the mechanisms responsible for controlling behavior in the task. In the next section, an ideal performer analysis is conducted to determine the optimal compensation strategy given the uncertainty inherent to the Bayes’ ball experiment.

### 4.4 Ideal performer analysis

The Bayes' Ball paradigm was designed to test whether humans are implicitly aware of a particular embodied dynamic, the Weber law of interval timing, and are able to exploit this awareness to improve their performance in the task. This was accomplished by presenting interval estimation under risk in a simple game-like task. During the test phase of the experiment, subjects were required to time their responses such that the ball stopped in one of three point regions. Faced with uncertainty about the exact interval of time that has elapsed since the ball moved behind the occluder, the optimal strategy in Bayes' ball was not to aim for the center of a target, but rather to aim for a location shifted slightly in the direction away from neighboring penalty regions. The magnitude of this compensation should be sensitive to both the uncertainty in the estimated interval, as well as the costs and gains associated with the different possible outcomes. An ideal
performer was used to obtain the optimal decision strategy, and determine whether human behavior conformed to the predicted optimal performance.

In particular, if subjects have an internal estimate, \( \tau \), of the amount of time that has passed since the ball disappeared from view, then according to the Weber law of timing, this estimated time is related to the physical elapsed time \( t \) according to a probability distribution \( p(t|\tau) \) with standard deviation increasing linearly with \( \tau \). In the simplest case, we assume a Gaussian distribution with mean \( \tau \) and standard deviation \( \sigma_0 + \sigma_1 \cdot \tau \):

\[
p(t|\tau) = \frac{e^{-\frac{(t-\tau)^2}{2(\sigma_0 + \sigma_1 \cdot \tau)^2}}}{(\sigma_0 + \sigma_1 \cdot \tau)^\sqrt{2\pi}}.
\]

Assuming the ball's velocity is known with certainty\(^{10}\), then the location of the ball given physical time \( t \) is \( p(x|t) = \delta(x - v \cdot t) \), where \( \delta \) is the Dirac delta function:

\[
\delta(z) = \begin{cases} 
1 & z = 0 \\
0 & z \neq 0
\end{cases}.
\]

From these two distributions, it is shown in Appendix B that the distribution of the ball's position given the internal time estimate is given by

\[
p(x|\tau) = \int_{-\infty}^{\infty} p(x|t) p(t|\tau) dt
\]

\[
= \frac{e^{-\frac{(x-v\tau)^2}{2v^2(\sigma_0 + \sigma_1 \cdot \tau)^2}}}{v(\sigma_0 + \sigma_1 \cdot \tau)^\sqrt{2\pi}}.
\]

Facing a set of target regions with different point values, the Bayesian optimal decision strategy is to stop the ball at subjective time \( \tau \) maximizing expected value:

\[
U(\tau) = \int_{x} u(x) p(x|\tau) dx,
\]

\(^{10}\) This simplifying assumption is made to keep the equations in a tractable form. In theory, after the practice phase of the experiment a Bayesian observer would have an accurate estimate of velocity, since it is constant on each trial. In practice, any residual uncertainty in velocity is absorbed in fitting the parameters \( \sigma_0 \) and \( \sigma_1 \) to human data.
where the point value of stopping at a location $x$ is given by $u(x)$. In the particular case of Bayes’ ball, the utility function $u(x)$ is directly specified to the subject in the task instructions as the point value associated with each region. An analytical form for $U(\tau)$ is provided in Appendix B. In the first experiment, if the ball reached the right side of the screen then this was counted as a penalty. Thus, there were effectively four point regions on each trial (the three visible regions, and the ‘invisible’ penalty region beginning at the right edge of the display). If the boundaries between regions are given by $x_1$, $x_2$, and $x_3$, and the reward values associated with the four regions are $r_1$, $r_2$, $r_3$, and $r_4$, then

$$u(x) = \begin{cases} r_1, & x \leq x_1 \\ r_2, & x_1 < x \leq x_2 \\ r_3, & x_2 < x < x_3 \\ r_4, & x_3 \leq x \end{cases}$$

The optimal decision time, $\tau_{opt}$, is the value maximizing $U(\tau)$ and is a function of the subject's internal estimation dynamics, the distance to the target on the current trial, as well as the configuration of point regions on the each trial. Presumably, while humans cannot eliminate the uncertainty in their temporal estimation, given opportunity and motivation they should be able to compensate for it by adjusting their intended stopping position closer or farther from the penalty zones.

In order to compare human and ideal performance, the parameters $\sigma_0$ and $\sigma_1$ were fit to each subject’s data. This procedure was carried out as follows. Data used for fitting the ideal performer models was restricted to both-penalty trials, since in this condition it can be assumed with some confidence that subjects were aiming for a relatively fixed point on each trial. By contrast, in penalty-before and penalty-after trials, it is possible that subjects had a ‘region of indifference’, for example, a strategy of “aiming anywhere on the left half of the display”. For each target distance, the empirical human data can be modeled as samples drawn from a particular distribution, namely $p(x_{obs}|\tau_{aim}, \sigma_0, \sigma_1)$, where $\tau_{aim}$ is the subject’s time estimate corresponding to their intended aiming position. The values $\sigma_0$ and $\sigma_1$ correspond to the dynamics of temporal uncertainty for that
subject. Since there were three different target distances, each subject had three different values of $\tau_{aim}^d$, one for each target distance $d \in \{ \text{near,middle, far} \}$. The parameters $\sigma_0$ and $\sigma_1$ relate to the subject’s inherent ability to estimate time, and are invariant with respect to the target distance. Thus, for each subject, five parameters characterized their distribution of stopping positions for the both-penalty trials at each of the three target distances. The values for these parameters were then chosen for each subject that maximized the likelihood of that subject’s empirical data:

$$\prod_{d \in D} \prod_{i=1}^{N_d} p(x_i^d | \tau^d, \sigma_0, \sigma_1),$$

where $x_i^d$ is the recorded ball position on the $i^{th}$ trial of the human data at target distance $d$. In implementation, the negative of the summed log-likelihood was minimized rather than maximizing the likelihood function\(^{11}\),

$$-\sum_{d \in D} \sum_{i=1}^{N_d} \ln p(x_i^d | \tau^d, \sigma_0, \sigma_1).$$

In fitting the ideal performer models, it was discovered that the human data included outliers that could not adequately be captured by a Gaussian distribution. To deal with this, outliers differing by more than 2 standard deviations from the mean for each configuration were excluded. This was done as maximum likelihood procedures are known to be sensitive to outliers that violate the assumptions of the underlying distribution (Ratcliff, 1993). In summary, the ideal performer model was obtained by fitting five parameters for each subject. However, note that three of the parameters, $\tau^d$,

---

\(^{11}\) Minimizing the negative log-likelihood is equivalent to maximizing the likelihood function, as the extrema of the log-likelihood function occurs at the same location as the likelihood function (since logarithm is a monotonic transformation). The logarithm of the likelihood function is typically used in practice as it allows working with sums rather than products and avoids numerical instabilities associated with multiplying small floating point values. The minimization was performed using the unconstrained nonlinear optimization routines in Mathematica (Wolfram Research, 2007).
For the results of the ideal performer analysis to be meaningful, it must be the case that the models accurately capture the constraints on performance for each participant. As one test of this, the predicted total scores were computed for each participant using the parameters $\sigma_0$, $\sigma_1$, and $\tau^d$. The question is whether the model, adopting the same decision strategy as the human subject, would achieve the same total score on the experiment. If that were the case, it would demonstrate that the performance constraints on the model (namely $\sigma_0$ and $\sigma_1$) accurately reflect those facing the human participant. A paired $t$-test comparing predicted and observed total scores found no significant difference ($p > .05$) between the predicted and empirical scores. Thus, by adopting the decision strategy used by humans, the model obtained the same resulting score.

Once the parameters $\sigma_0$ and $\sigma_1$ were obtained for each subject and the validity of the model constraints verified, the $\tau^d$ parameters were discarded, since the objective of the ideal performer analysis is to predict what humans optimally should do, rather than empirically fit what they did do. Optimal performance was predicted for each subject by numerically computing $\tau_{opt}$, or the stopping criterion that maximized the expected value $U(\tau)$. The observed human data was then compared to the model's prediction, given by the distribution $p(x|\tau_{opt}, \sigma_0, \sigma_1)$.

Figure 4-8 compares the optimal distribution of stopping positions, $p(x|\tau_{opt})$ to actual human performance for a single subject on the distance far, penalty-after trial configuration. The histogram (blue bars) displays the observed human performance, while the smooth curve in the top panel shows the expected performance of an optimal decision-maker constrained to the estimation uncertainty of the human subject. The reward and penalty configuration is overlaid at the top of the figure. In this example, both the human and ideal performer have shifted their mean stopping position away from the penalty region, although the magnitude of the shift is slightly smaller for the human than for the ideal performer. The bottom panel shows the expected mean score as a function of intended mean aiming position. The maximum of this curve corresponds to
the aiming position that would maximize the subject’s score on this target distance and configuration of penalty regions.

**Figure 4-8.** Comparison of performance of a single subject to an ideal performer calibrated to that subject. (Top) The smooth curve shows the predicted optimal stopping point distribution, while the histogram (blue bars) shows the empirically observed distribution. (Bottom) The dashed curve shows the expected value (in points) as a function of mean stopping position. The maximum of the dashed curve corresponds to $U(\tau_{opt})$. The marked point shows the mean stopping position and resulting score for the human subject’s data shown in the histogram at top. The error bars (vertical and horizontal) indicate 95% confidence intervals in mean position ($x$ error bars) and score ($y$ error bars).

For each subject, the ideal performer model was used to compute the optimal mean stopping position for each distance and penalty configuration. Table 4-2 compares the optimal strategy to the empirically observed strategy. One surprising feature of the data in Table 4-2 is that the optimal mean stopping position on both-penalty trials does not correspond to the exact center of the reward region. Instead, the optimal strategy is to undershoot the center of the reward region, by 0.72 mm for the near targets, and 1.72 mm for the far targets. A similar undershoot bias was observed in the human data, though for the human subjects the magnitude of this bias is much greater. In the case of the ideal performer models, the bias stems from the fact that uncertainty increases with increasing distance, not just across large changes in target distance, but also for relatively small differences in aiming position for a given target distance. By aiming slightly to the left of the center of the reward region, the ideal performer can slightly
reduce its variability, and thus increase its expected score higher than what it can achieve by aiming for the center of the target.

**Table 4-2. Mean target-relative stopping position (delta) and score for human subjects and the ideal performer analysis.**

<table>
<thead>
<tr>
<th>Distance</th>
<th>Config</th>
<th>Delta</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Empirical</td>
<td>Optimal</td>
</tr>
<tr>
<td>Near</td>
<td>Penalty-before</td>
<td>6.1</td>
<td>3.37</td>
</tr>
<tr>
<td></td>
<td>Both-penalty</td>
<td>3.01</td>
<td>-0.72</td>
</tr>
<tr>
<td></td>
<td>Penalty-after</td>
<td>0.62</td>
<td>-4.15</td>
</tr>
<tr>
<td>Middle</td>
<td>Penalty-before</td>
<td>4.35</td>
<td>8.65</td>
</tr>
<tr>
<td></td>
<td>Both-penalty</td>
<td>-0.76</td>
<td>-1.09</td>
</tr>
<tr>
<td></td>
<td>Penalty-after</td>
<td>-3.98</td>
<td>-8.36</td>
</tr>
<tr>
<td>Far</td>
<td>Penalty-before</td>
<td>1.79</td>
<td>12.75</td>
</tr>
<tr>
<td></td>
<td>Both-penalty</td>
<td>-3.07</td>
<td>-1.72</td>
</tr>
<tr>
<td></td>
<td>Penalty-after</td>
<td>-12.36</td>
<td>-18.14</td>
</tr>
</tbody>
</table>

A measure of compensation was computed for the ideal performers in exactly the same manner as was done for the human data, by subtracting stopping position on both-penalty trials from stopping position on penalty-before and penalty-after trials, and changing the sign for penalty-after trials such that positive values of compensation consistently indicate shifting away from the penalty region. Compensation is plotted for human participants and the ideal performers in Figure 4-9. It is immediately apparent that while human subjects demonstrated appropriate compensation due to the asymmetrical penalty configurations, the magnitude of their compensation was well below the Bayesian optimal decision strategy. Interestingly, for both human subjects and the ideal performer model, compensation was greater on penalty-before trials than penalty-after trials, with the exception of the far targets, where this pattern was reversed. On the far targets, both the humans and the model exhibited greater compensation for the penalty-after trials than the penalty-before trials. In the case of the ideal performer model, this crossover is due to the penalty associated with reaching the right edge of the display, causing the ideal performer to decrease its compensation for penalty-before trials at the far target distance.
Figure 4-9. Compensation relative to mean stopping position on both-penalty trials for each target distance. Solid lines reflect empirical data, while the dashed lines are the ideal performers. Error bars on the human data indicate 95% confidence intervals.

Human subjects clearly adapted their behavior to the dynamics of uncertainty in interval estimation and were able to exploit this knowledge to improve performance, but based on the ideal performer analysis they failed to do so in a perfectly optimal manner. Unfortunately, it is not clear if the failure stems from imperfect knowledge of their variability, imperfect ability to combine this knowledge with a task-specific utility structure, or some other deficit in exploiting all of the available information. Another possibility is that subjects were biased \textit{a priori} towards aiming for the center of the targets, and the cost structure of the task environment provided insufficient incentive for subjects to compensate more than they did. Such an explanation is plausible, as in most natural environments a target is something that is \textit{aimed for}, rather than \textit{away from}. In addition, during the practice phase (in the absence of penalty regions) subjects were explicitly told to aim for the center of the target.

In order to assess the cost associated with subjects’ deviation from the optimal compensation strategy, the mean score for each human subject was compared to the expected score obtained by the corresponding ideal performer model. Across subjects, the mean score difference was -8.96 points (SD = 5.11). Thus, despite their deviation from optimal performance in terms of stopping position, the difference in terms of score was small relative to the potential outcomes of each trial (-200, 0, or +100 points), suggesting that human performance was not that far removed from the optimum.
4.5 Cognitive models of Bayes’ Ball

The behavioral data from the Bayes’ Ball experiment supports the hypothesis that subjects acquired and exploited an internal model of the embodied dynamics of the task. The ideal performer analysis indicated that subject performance was nearly optimal, and thus provides support that subjects in fact acquired internal models of their variability in estimating time intervals.

A limitation of the ideal performer analysis is that it assumed omniscience on the part of the subjects. Subjects were assumed to come in to the experiment already knowing with maximal confidence the limitations of their ability to estimate time intervals, as well as the exact velocity of the ball on each trial.

A claim of this dissertation is that adaptation to the embodied dynamics of routine interactive tasks is achieved through the acquisition of internal models on the basis of experience in a task. It remains to be addressed, therefore, whether the observed behavioral results could be explained by a theory that must learn from its (limited) experience. At the same time, it remains to be seen whether the same pattern of results could be explained by a theory that does not involve the use of internal models. The purpose of this section is to address these two issues. Cognitive modeling can be used as a formal specification of a verbal theory, but more importantly as a sufficiency test that the theory is in principle capable of accounting for the observed performance. Here, two different cognitive models will be presented. Both models learned to perform the task, starting only with the background knowledge and task instructions that human subjects had when beginning the task.

The two models presented here differ in terms of the cognitive representations acquired and exploited during performance of the task. The first model explicitly tests the hypotheses of this thesis, by learning an internal forward model of the embodied dynamics of estimating short time intervals. This model is implemented using the adaptive internal model framework, developed and described in Section 3.4.1. The second model acquires only a representation of the reward structure of the task, using reinforcement learning with a feedforward neural network for function approximation (Sutton & Barto, 1998). In the latter case, it might be argued that the cognitive model, through its use of a neural network, is still acquiring an internal model of the reward
structure of the task. Regardless, it will be shown that the reinforcement learning model is unable to account for the timescale of human learning in the experiment.

4.5.1 Time estimation and constraints on performance

At the core of the Bayes’ Ball task is the ability to estimate the passage of short time intervals. The crucial action (and source of uncertainty) in the Bayes’ Ball experiment is not a motor act, but rather the cognitive act of waiting for a certain temporal interval before releasing the response key. To enable the cognitive models to perform the task, both models were endowed with the ability to estimate time intervals, using a psychological theory of prospective time estimation as implemented by (Taatgen et al., 2007). This particular model of time perception is based on the idea of an internal clock that stochastically generates pulses. By accumulating pulses, the cognitive system is able to estimate the passage of time. The clock mechanism generates pulses with the following dynamics: The time until the occurrence of the first pulse is given by a parameter $t_0 = \text{startpulse}$. Each pulse is separated from the previous by an interval that is some constant multiplier $a > 1$ of the previous interval. The exact duration of each interval is a stochastic variable, using noise drawn from a logistic distribution with zero mean and standard deviation dependent on the magnitude of the current interval. This yields the recurrence relation

$$
t_0 = \text{startpulse},
$$

$$
t_{n+1} = a \cdot t_n + \text{noise}(\mu = 0, \sigma = b \cdot a \cdot t_n).
$$

The parameters startpulse, $a$, and $b$ were all kept at the values reported in (Taatgen et al., 2007).

In addition to their ability to estimate elapsed time, the cognitive models were restricted by motor variability. The time required to execute the motor act of releasing the response key was modeled as a Gaussian random variable, with a mean of 100 ms and standard deviation of 25 ms. The value of 100 ms was based on empirical estimates of the time required to release a mouse button (John, Vera, Matessa, Freed, & Remington, 2002). The standard deviation for motor execution variability was chosen somewhat arbitrarily, but is roughly consistent with human performance and was not varied to fit human data.
The models interacted with a simulated version of the task that maintained the core functional properties of the human experiment, such as the target distances and ball velocity. In addition, the models faced the same organization of the experiment: practice trials with no point score, followed by test trials grouped into blocks of nine trials, with each combination of distance and configuration occurring once in each block.

4.5.2 Cognitive model based on the adaptive internal model framework

A cognitive model based on the adaptive internal model framework (Section 3.4.1) was developed to learn to perform the Bayes’ Ball task. At the core of the model is the internal model used to generate predictions regarding the embodied dynamics of the task. Unlike previous theories of internal models in motor control, the crucial action in the Bayes’ Ball environment is not a motor act, but rather the cognitive act of waiting for a certain temporal interval before releasing a response key. The internal model took the form of a feedforward neural network with one input unit, corresponding to a number of pulses accumulated from the internal clock. The input values were normalized to lie in the range \((0,1)\). There were two output units of the neural network. One corresponded to the predicted mean distance that the ball would travel in a given time interval. The other output unit corresponded to the predicted standard deviation in the ball’s position, assuming Gaussian variability. Thus, the neural network learned an internal model of the probability distribution of the ball’s location given the number of pulses that have elapsed, \(p(x|\tau) \sim N(\mu(\tau),\sigma(\tau))\), where both the distribution’s mean and standard deviation were arbitrary nonlinear functions of the number of pulses. The neural network’s input unit and mean output units used linear transformation functions, while the standard deviation output used an exponential transformation, \(\phi(x) = e^x\). This was done to ensure that the network predicted only positive values of standard deviation. The network used 25 hidden units with tanh transformation functions.
Figure 4-10. Neural network structure employed for the internal model of embodied dynamics learned in the Bayes’ Ball experiment.

Figure 4-10 shows a graphical representation of the structure of the neural network. The number of hidden units was deliberately chosen to be larger than necessary to learn the task. To avoid overfitting the network to the data, the input data for the network was split into a training and validation set. The network was trained for an overly large number of iterations (500), and the network weights resulting in the lowest error on the validation set were used as the learned network parameters. As the neural network was learning a probabilistic model, a maximum likelihood error function was used rather than minimizing sum-of-squares error. In particular, the negative log-likelihood of observing a set of resulting ball positions \( \{x_i\}, \; i = 1 \ldots N \) given a set of perceived elapsed intervals \( \{\tau_i\}, \; i = 1 \ldots N \) is given by

\[
E = -\sum_{i=1}^{N} \ln \left\{ p(x_i|\mu(\tau_i,w),\sigma(\tau_i,w)) \right\},
\]

where \( p(x|\mu,\sigma) \) is normally distributed with mean \( \mu \) and standard deviation \( \sigma \), and \( \mu(\tau,w) \) and \( \sigma(\tau,w) \) are the output units of the neural network, as a function of the network inputs and the adaptive network weight parameters \( w \). The derivative of this error function was evaluated with respect to each of the network outputs and

---

\(^{12}\) The use of a validation set could have been avoided by limiting the number of hidden units, limiting the amount of training, or introducing a regularization term into the training procedure (Bishop, 2006). However, each of these approaches would have resulted in the introduction of additional free parameters.
backpropagated through the network to adjust the weight parameters. The optimization of the weights was performed using the BFGS algorithm (Nocedal & Wright, 1999).

For a given input, the internal model produced a normal distribution describing the ball’s predicted position. The immediate output of the neural network was a predicted mean and standard deviation. However, rather than using these statistics as the final output of the internal model, the model predictions were instead represented as a set of random samples drawn from the distribution defined by the given mean and standard deviation. The reason for producing output in terms of samples rather than sufficient statistics (mean and standard deviation) was twofold. First, this allowed the predictions of the forward model to use the same representational form as actual experience acquired through interaction with the task (i.e., samples rather than statistics). Second, individual outcomes can be directly evaluated in terms of their utility: both human subjects and the model were instructed at the beginning of the experiment that if the ball stopped in the green region the resulting score was +100 points. In contrast, a predicted mean and standard deviation cannot be easily assigned a utility value. As shown by the ideal performer analysis, the mathematical relationship between standard deviation and expected score is nontrivial. The number of samples generated for each input was chosen arbitrarily to be 10, and this value was not adjusted to fit human data.

In addition to the forward model, the cognitive model also contained several other interacting modules that enabled it to perform the task. Each of these modules are part of the adaptive internal model framework previously described in Section 3.4.1 and Figure 3-8:

- A memory module stored actions and outcomes. Actions were defined as the temporal intervals that the agent waited before releasing the response key (measured in internal clock pulses). Outcomes were the resulting ball positions (note that the score feedback given to subjects was not used by this model).
- An action generation module was used to generate a set of candidate actions on each trial. In the Bayes’ Ball model, these actions were simply pulse counts sampled uniformly from the range \((0, 40)\). According to the internal clock theory, this range corresponded to physical time intervals of between
roughly 0 and 6 seconds, a range much larger than needed to perform the task. Pulse counts were rescaled to lie in the range \((0,1)\) for input to the neural network. The number of candidate actions generated for each trial was chosen arbitrarily to be 100, this value was not adjusted to fit human data. No learning was modeled in the action generation module.

- A reward module took as input ball positions (either predicted or experienced) and a trial configuration, and produced as output the utility values associated with those ball positions. During the practice phase, it was assumed that subjects attempted to minimize the squared error between the ball position and target center. Thus, the reward function employed was 

\[ r_{\text{practice}} = -\left( x_{\text{ball}} - x_{\text{target}} \right)^2. \]

During the test phase, it was assumed that subjects attempted to maximize their point score, and the reward function was simply the point values associated with stopping in the different regions. For example, the reward for stopping in the green region was +100. Learning of the reward function was not modeled, as it was assumed that the human subjects learned to evaluate trial outcomes based on the verbal instructions given to them before the start of the task.

- The model also contained an action selection module. This module took as input a set of actions (produced by the action generation module) and for each action a random sample of resulting utility values (produced by the combination of the internal model and reward modules). The action selection module accumulated the utility for each action and simply selected the action that had the highest accumulated value. No stochasticity was modeled in the action selection module.

- Finally, the model contained a crude sensorimotor system. This system simulated pressing a key to start each trial, and after the intended number of pulses had elapsed, released the key. The model was able to perceive the target distance and penalty configuration on each trial, and the stopping position of the ball. The unit of distance was arbitrarily chosen to be pixels;
equivalent performance could have been obtained using a more ecologically valid estimate such as degrees of visual angle.

Although the overall model contained a number of interacting components, the only learning component was the neural network in the internal model. The parameters in the model were either chosen arbitrarily and not adjusted to fit human data (in the case of the number of samples from the action generation module and internal model), or estimated based on prior research (in the case of the motor execution time and the parameters for the internal clock).

Twenty-three simulated subjects performed the Bayes’ Ball experiment. This number of subjects was chosen for the sake of computational tractability. Figure 4-11 shows the average learning performance of the human subjects and cognitive models across practice blocks of the experiment. Performance is measured in terms of the mean absolute difference between the ball position and the center of the target. Each practice block consisted of one trial at each of the three target distances. As can be seen from the figure, the cognitive model quickly learned to perform the task and achieve low error. The asymptotic performance of the model was limited by the uncertainty inherent in its temporal estimates and the motor noise involved in releasing the response key. This performance closely matched human subject performance, despite having no free parameters in the model.
Figure 4-11. Learning curves for human subjects and the adaptive internal model architecture across practice blocks of the experiment. The curves represent mean absolute difference between stopping position and the target center. Each block consists of three trials (one trial at each target distance).

The model did appear to perform worse than human subjects in the first few trials. Humans achieved near asymptotic performance within roughly 5 practice blocks (or 15 trials). For the model, it took closer to 15 blocks to achieve asymptotic performance. This difference might plausibly be explained by the background experience that human subjects brought in to the experiment. A Bayesian framework (Weiss et al., 2002) has been shown capable of accounting for a range of optical illusions by assuming that humans possess a prior belief distribution over physical velocities. In the Bayes’ Ball experiment, such knowledge could have benefited human subjects on the early trials, whereas the cognitive model (as implemented) had no such prior knowledge. Similarly, human subjects were told that the ball moved at constant velocity, whereas the cognitive model made no assumptions about the physics of the environment with which it interacted.

Figure 4-12 illustrates the acquired internal model for one of the simulated subjects at the end of the practice trials. The top graph shows the internal model’s predicted mean ball position as a function of the subjective elapsed time interval (black curve). The red curves show the predicted uncertainty in the ball’s position in terms of 95% confidence intervals (computed as $\pm 1.645 \times \sigma$). The black markers in the graph indicate the experienced ball dynamics (the outcomes of previous trials) on which the internal model
was based. Although the ball moves at constant velocity, the functional relationship between subjective elapsed time (measured in clock pulses) and ball position is nonlinear due to how time is encoded by the internal clock theory. Figure 4-12b shows just the predicted standard deviation as a function of clock pulses. The Weber law predicts a linear increase in standard deviation with increasing physical time, but as with the mean ball position, the relationship acquired by the model is nonlinear due to the nonlinear encoding of subjective elapsed time.

Figure 4-12. Internal model of embodied dynamics in the Bayes’ Ball experiment. (a) The predicted mean ball position (black curve) as a function of subjective elapsed time, with 95% confidence intervals overlaid (red curves). The marker points indicate the model experience from previous trials. (b) The predicted standard deviation in ball position as a function of subjective elapsed time. Although the Weber law predicts a linear increase in standard deviation with
increasing time, the nonlinear temporal encoding of the internal clock predicts a nonlinear relationship between pulses and uncertainty.

Learning was turned off for the cognitive models at the end of the practice trials, as it was observed that all of the models acquired an accurate internal model of the embodied dynamics of the task well before the end of the practice phase. This was done simply to minimize the amount of computation time required to run the simulations\textsuperscript{13}. The cognitive models were then run on the test phase of the experiment. Although the cognitive models did not learn anything beyond the internal model acquired during the practice trials, they were able to closely match human performance during the test trials of the experiment. Figure 4-13 shows the mean scores for human and model subjects across four epochs of the test trials (each epoch consisted of 225 trials). The model’s mean score was within the 95\% confidence intervals of the human performance for all four epochs, although human subjects did exhibit some learning from the beginning to the end of the experiment.

\textbf{Figure 4-13. A comparison of mean score across four epochs of the test trials for human subjects (solid markers) and model simulations (open markers).}

The decision strategy adopted by the cognitive models was analyzed by computing the difference between the mean stopping position and target center for each target distance and penalty configuration (equivalent to the delta measure computed for human

\textsuperscript{13} Training twenty-three simulated subjects for just the practice phase of the experiment required roughly 12 hours of computation time on an 2.33 GHz Intel MacBook Pro.
subjects). These results are shown in Figure 4-14, top row. Compared to the human subjects, the cognitive models exhibited much greater compensation away from the penalty regions (approximately twice as much), and as such the cognitive models’ performance was much closer to the predictions based on the ideal performer analysis. In addition, the overall bias observed in subject data—overshooting near targets, and undershooting far targets—was not observed in the model’s behavior. Despite these differences in behavior, the model and human subjects achieved very similar scores in the task, as evidenced by Figure 4-13. This finding suggests that humans were in fact nearly optimal in the task according to the performance objective they were given (maximizing their score).

Figure 4-14. Comparison of human and model performance on the test trials of Bayes’ Ball experiment 1. The top row compares the human data (left graph) with the parameter free model (right graph). The bottom row compares human data (repeated from the top row) with a model incorporating two biases. See text for details. For the bottom two panels, $r^2 = 0.95$. 
In order to investigate the possible reason for the biases observed in the human data, the cognitive model was subsequently modified in two ways. First, it was hypothesized that human subjects might be biased towards selecting aiming points near the center of the reward region, as this is the task the human subjects explicitly adopted during the practice phase of the experiment. To account for this target-center bias, the reward module was modified to value stopping closer to the target more highly. In particular, the modified reward function used was $r_{\text{total}} = r_{\text{test-trial}} + \alpha \cdot r_{\text{practice-trial}}$, where $r_{\text{test-trial}}$ represents the objective point value associated with a particular outcome, $r_{\text{practice-trial}}$ is the reward function that was used during the practice phase of the experiment (based on minimizing squared error from the target center), and $\alpha$ was a weight parameter, chosen to be 0.006.

The other bias observed in the human data was a general tendency to overshoot near targets and undershoot far targets. To account for this bias, it was assumed that the cognitive model incorrectly estimated the target distance on each trial. In particular, the estimated target distance was computed as $x_{\text{estimated}} = (1 - \beta) \cdot x_{\text{target}} + \beta \cdot \bar{x}_{\text{target}}$, where $x_{\text{target}}$ was the actual target distance, $\bar{x}_{\text{target}}$ was the mean estimated target distance from all previous trials in the experiment, and $\beta$ was a weight parameter set to 0.04. A value of $\beta = 0$ would reflect zero bias in distance estimation. The strategy of basing a perceptual estimate such as a distance judgment on both current and previous perceptual information is rational assuming that the perceptual estimate on any given trial is uncertain or corrupted by noise and that the world is unchanging. Such an estimation strategy leads to biased estimates in the Bayes’ Ball environment due to the randomized order of trial distances in the experiment. Such biases towards the mean of previous stimuli are commonly observed in psychophysical experiments (Gescheider, 1988; Mozer, Kinoshita, & Shettel, 2007). The simulations were re-run using these two biases ($\alpha$ and $\beta$), and the results are shown in the bottom row of Figure 4-14. With only these two alterations, the model was able to closely reproduce the human data ($r^2 = 0.95$).
4.5.3 Reinforcement learning model

The previous section demonstrated the explanatory sufficiency of the main theory of this dissertation with regard to the empirical data observed in the Bayes’ Ball task environment. This theory states that humans can acquire and exploit internal models of their embodied dynamics in order to control and improve their performance in routine interactive tasks. The particular dynamic relevant to the Bayes’ Ball experiment is not a perceptual or motor dynamic, but rather the cognitive ability to estimate short temporal intervals. Although the sufficiency of this theory was demonstrated by implementing the theory in the form of a computational cognitive model, it remains to be seen whether an alternative account can also account for the same pattern of results observed in the human data. Namely, can a cognitive model that doesn’t acquire or exploit an internal model of temporal estimation uncertainty learn to perform the task as quickly and as well as human subjects?

This section will demonstrate that a model based on acquiring a mapping from actions directly to their task-specific utilities (using reinforcement learning) is unable to account for human performance in this task. Reinforcement learning is a very general approach in machine learning, perhaps best defined in terms of how it differs from supervised learning. In supervised learning, the agent learns by attempting to minimize the discrepancy between its computed and desired outputs. This learning scheme therefore requires the availability of what the desired outputs should be. In the terminology of supervised learning, these desired outputs are provided by an external teacher. Note that the ‘teacher’ can be rather abstract, and may be constituted by the environment itself. In many scenarios, such a teaching signal is not available and only a scalar reinforcement signal is available to the agent. This reinforcement signal indicates whether the agent’s output is generally good or bad, but does not indicate how it needs to adjust its output to improve performance. Just as supervised learning is not a single algorithm but rather a learning framework, there is no single reinforcement learning algorithm. A thorough review of reinforcement learning methods is given in (Sutton & Barto, 1998).

For the purpose of this section, there are two properties of reinforcement learning that make it an interesting and relevant counter-hypothesis to the internal model theory
of embodied cognition. First, under certain restrictive circumstances, particular reinforcement learning algorithms such as Q-learning can be shown to be guaranteed to learn an optimal behavioral policy (Watkins & Dayan, 1992). This property suggests that despite the impoverished information used as a teaching signal, reinforcement learning can be a powerful computational learning framework. Second, reinforcement learning algorithms can learn to perform optimally in a task without acquiring any internal models of the environment or its dynamics. If a model-free learning algorithm can account for human performance in the Bayes’ Ball experiment, then it would seriously weaken the hypothesis that human performance is based on acquiring such internal models. Note that strictly speaking, reinforcement learning does in fact acquire an internal model of the reward structure of the environment. Similarly, there are numerous variations on reinforcement learning that do explicitly incorporate internal models of the environment or the consequences of actions (Sutton & Barto, 1998), and so the distinction between internal models and reinforcement learning is not intended to be mutually exclusive.

Reinforcement learning has previously been used to model the acquisition of human motor skill. In particular, Berthier and colleagues (Berthier, 1996; Berthier, Rosenstein, & Barto, 2005) simulated the development of reaching ability in infants using model-free reinforcement learning. In their model, achieving near asymptotic performance required on the order of 30,000 reaching trials. This amount of experience may in fact be a realistic assumption for the time course of motor learning in infants, but it also highlights a general challenge of the reinforcement learning framework. By limiting access to or ignoring all environmental feedback apart from a scalar reward signal, the learning algorithm faces the challenge of discovering optimal performance in a vast space of possible strategies given greatly impoverished feedback. When the outcomes of actions are available in addition to their reward values, learning speed can be greatly improved. Indeed, (Berthier et al., 2005) acknowledge that the reinforcement learning framework they propose cannot account for some aspects of infant motor learning that would appear to require a forward model of arm dynamics.
Like the embodied dynamics model, at the core of the reinforcement learning model is a neural network. Whereas in the embodied dynamics model this neural network was used to learn the mapping between actions and their outcomes in the task, the reinforcement learning model attempted to directly learn a mapping between actions and their task-specific utility values, eliminating the necessity of an internal model of the Weber law.

It was shown in the previous section that the embodied dynamics model learned to perform the task near-optimally on the basis of the practice trials of the experiment. In fact, learning was disabled before the start of the test trials. In contrast, reinforcement learning can’t, even in theory, learn to perform near-optimally by the start of the test trials, as a reinforcement signal (the point score resulting from each action) is not given until after the first trial of the test phase. Therefore, the critical test of the reinforcement learning model is how quickly it can learn once the test trial phase of the experiment begins.

Since the utility of an action depends on both the target distance and penalty configurations, the neural network used for the reinforcement learning model included additional input units encoding these variables. Actions were encoded as before, by scaling the number of elapsed internal clock pulses to lie in the range $(0,1)$. The target distance was similarly represented by a single numerical value in the range $(0,1)$. The penalty configuration, which is a categorical variable, was represented using a 1-of-$n$ coding scheme that mapped the penalty configuration onto three input units:

- penalty-before $\rightarrow \langle 1,0,0 \rangle$
- both-penalty $\rightarrow \langle 0,1,0 \rangle$
- penalty-after $\rightarrow \langle 0,0,1 \rangle$.

This is a standard approach to representing categorical variables in neural networks. Thus in total, the neural network used five input units: 1 action unit, 1 distance unit, and

\[ \text{penalty-before} \rightarrow \langle 1,0,0 \rangle, \]
\[ \text{both-penalty} \rightarrow \langle 0,1,0 \rangle, \]
\[ \text{penalty-after} \rightarrow \langle 0,0,1 \rangle. \]

\[ \text{14} \] A neural network was necessary as the action space is continuous, and therefore some form of function approximation is necessary. A neural network was used in particular (as opposed to an alternative function approximation scheme) in order to minimize the differences between the embodied dynamics and reinforcement learning models.
3 penalty configuration units. Twenty-five tanh hidden units were used, and a single linear output unit was used to predict the expected utility.

The reinforcement learning model was trained as follows. A training set was generated by executing random actions in the test trial phase of the experiment, and observing their resulting utility values. This resulted in a database of experience. The reinforcement learning model was evaluated by examining task its performance as a function of the amount of prior experience it was given. The amount of experience was varied from 0 to 900 trials in six increments. The reason for using this training and evaluation scheme was simply for computational tractability—using this approach, a single simulated subject only had to be trained six times (once at each level of prior experience). By contrast, running a simulated subject through the entire experiment would require retraining the neural network 900 times, on an ever-increasing training set size.

For each experience level, the database of experience was divided into training and validation sets, which were used to train the neural network (as before, using the BFGS optimization algorithm). The performance of the trained network was then evaluated by turning off learning, and completing 900 additional test trials of the experiment. Actions were chosen by selecting the action that the reinforcement learning model predicted had the highest expected value. A learning curve was then computed in terms of the resulting mean score as a function of the amount of training that the neural network had been given.

Since this training and evaluation procedure differed than the one used to evaluate the adaptive internal model framework, a modified version of that model was also retrained and evaluated using the same procedure. The neural network used for the modified adaptive internal model framework was also extended to include distance and penalty-configuration input units. Since the embodied dynamics (the ball velocity and the agent’s temporal estimation uncertainty) did not vary as a function of target distance or penalty configuration, these features represented irrelevant, distractor inputs that the network would have to learn to ignore. This procedure eliminated differences in neural network architecture, training or evaluation methods as potential confounds.
The two model types were compared by simulating 15 agents for each model type, varying the number of training blocks from 0 to 100 in six increments (a single training block consisted of 9 trials—one at each distance and penalty configuration). The performance of each model type is shown in Figure 4-15, compared to human performance. The reinforcement learning model performs much worse than humans, even after 900 training trials. In contrast, the embodied dynamics model based on the adaptive internal model framework starts out performing worse than humans, but managed to reach human-level performance within 10 blocks of the experiment. It should be noted that the failure of the reinforcement learning model does not stem from any theoretical limitations in neural networks or network training procedures, as the embodied dynamics model successfully learned the task using the same network structure and training procedure.

Instead, the limitation of the reinforcement learning model is entirely due to the paucity of information available to the agent to learn from. By learning exclusively on the basis of scalar reinforcement signals, and throwing away the richness of information available in the world in terms of the dynamics of the task, the reinforcement learning model was destined for failure at the outset. Learning by acquiring models of task dynamics is a sufficient explanation for human performance in the task; learning through trial-by-trial reinforcement is an insufficient explanation.

These results strongly suggest that human subjects learned to perform the Bayes’ Ball experiment by acquiring and exploiting rich internal models of their embodied dynamics, rather than by learning to maximize their score during the test trials using only a scalar reinforcement signal. Further, the results suggest that much of this learning occurred during the practice phase of the experiment, when subjects could not have learned an optimal compensation strategy based on reinforcement (since no reinforcement for compensating away from the targets was given during the practice phase).
4.6 Summary

This chapter described an empirical study, an ideal performer analysis, and a series of computational cognitive models. Each of these contributions was directed towards the single goal of providing evidence that humans are capable of acquiring internal models of their embodied dynamics, and that they exploit these internal models to shape and improve performance in routine interactive tasks. Unlike previous research demonstrating evidence for internal models in motor control, the key embodied dynamics of the Bayes’ Ball paradigm were not perceptual or motor dynamics, but rather the uncertainty inherent in the human ability to estimate short temporal intervals—the Weber law of interval timing. The Bayes’ Ball paradigm constituted an environment where, if subjects were able to predict the consequences of the Weber law, they could exploit this knowledge to improve their performance. This is indeed what was observed in the experiment: subjects compensated away from penalty regions, and increased the magnitude of their compensation as the target distance increased.

An ideal performer analysis was conducted to determine whether human performance matched an optimal decision strategy, given the performance constraints imposed by the Weber law and features of the task environment. It was found that
human subjects closely approximated optimal performance. However, compared to the ideal performer, human subjects were consistently biased towards the center of targets. In addition, human subjects exhibited an additional bias in terms of overshooting the closest targets, and undershooting the farther targets. Despite the presence of these two biases in the empirically observed decision strategy, it was shown that the overall performance of the human subjects differed from optimal performance by less than 10 points per trial. This difference is relatively small, especially given that the outcome of each trial was either -200, 0, or +100 points.

Computational cognitive models were developed as an explicit and quantitative test of the explanatory sufficiency of the main theory of this thesis. These models were based on the adaptive internal model framework developed in Section 3.4.1. It was shown that, with no free parameters, a cognitive model that acquires and exploits an internal model of the embodied dynamics of the task was able to closely match human learning and asymptotic performance. When the model was extended with two parameters, it was also able to offer a plausible explanation for the nature of the small deviation from optimal performance.

An alternative cognitive model was also developed and evaluated, one that used reinforcement learning. This model represented a contrasting view to the embodied dynamics hypothesis, as it explicitly learned to perform the task without developing any internal representation of its performance dynamics. Despite the best efforts of the modeler, the reinforcement learning model was unable to account for the timescale of human learning or asymptotic human performance. It was demonstrated that this deficit was not due to the underlying neural network architecture, training, or evaluation procedure used for the model, but rather the nature of its cognitive representation of the task. Combined, the results of the computational cognitive models offer a strong argument in favor of the human capacity to acquire and exploit internal models of embodied dynamics.
5. Bayes’ Ball experiment 2

5.1 Overview and hypotheses

The previous chapter, in combining empirical data, ideal performer analysis, and cognitive modeling, provided converging support for the main hypotheses of this dissertation. Namely, in the Bayes’ Ball experiment, it was shown that humans are able to acquire and exploit internal predictive models reflecting the dynamics of uncertainty in temporal estimation, and combine this knowledge with knowledge of the external task environment and desired outcomes in order to optimize their performance. At the same time, the empirical results from the first experiment raised several additional questions. This chapter reports the results of a second experiment designed to address these questions, using a modification of the Bayes’ Ball paradigm.

One purpose of the experiment was to replicate the main findings from the first study. However, a more important motivation for the experiment was uncovering the origin of the discrepancy between human performance and the predictions of the ideal performer analysis. Compared to the ideal performer, the human data from the first experiment exhibited two biases. Subjects adopted a response strategy that appeared biased towards the center of each reward region. In addition, subjects consistently overshot the near targets, and undershot the far targets. It was hypothesized that this latter bias may be due to sequential interference effects stemming from the randomized order of target distances in the experiment. This hypothesis is consistent with a wide range of sequential estimation biases observed in the psychophysical literature (Gescheider, 1988). The second experiment was designed to test this hypothesis, by adopting a blocked trial design rather than the randomized design previously used.

In addition, a between-subject manipulation was introduced in the second experiment to rule out the possibility that subjects were adapting to some idiosyncratic feature of the Bayes’ Ball paradigm, rather than to variability in estimating the ball’s position *per se*. This was achieved by introducing artificial noise in the ball’s position for one half of the subjects, while keeping all other features of the task constant. In the first experiment, individual differences in time estimation ability were weakly correlated with magnitude of compensation. Incorporating an artificial noise condition allows
direct manipulation of individual estimation ability and therefore allows a more powerful test of differences.

Cognitive models were used in the last experiment to argue in support of an additional claim, namely that subjects were acquiring internal predictive models of the Weber law, rather than learning on the basis of trial-by-trial feedback. In this second experiment, a blocked trial design was adopted to provide further support for this argument, by designing a situation where trial-by-trial score feedback would not be a viable explanation for human learning performance. This was achieved by having all subjects complete a both-penalty block during the first block of the experiment, and then examining subjects’ decision strategies on the first trial of an asymmetrical penalty block (penalty-before or penalty-after). Up until this trial in the experiment, subjects were reinforced only for aiming for the center of the reward region. If subjects compensated away from the center of the reward region on this trial, this would suggest that they had learned to predict the variance in the ball’s position, rather than simply learned a mapping from time intervals to expected reward.

Finally, the second experiment also introduced a post-experiment survey designed to test subjects’ explicit awareness of the optimal decision strategy in the task. It is claimed that acquiring and exploiting internal models of embodied dynamics reflects the normal, ongoing mode of routine human interaction, and as such does not require explicit awareness or a deliberate strategy. If this were in fact the case, then it might be expected that the subjects’ decision strategies, implicitly defined by their performance in the task, would outperform their explicit judgments of an optimal strategy in the post-experiment questionnaire.

5.2 Method

5.2.1 Participants

Forty-eight undergraduate students at Rensselaer Polytechnic Institute volunteered to participate in the experiment in exchange for course credit. All of the subjects were naïve to the purpose of the study, and none had participated in the first experiment. The subjects’ mean age was 18.7 years (SD = 1.1). Eight of the subjects were female.
5.2.2 Apparatus and Materials

The apparatus used in this experiment was identical to that of the first experiment, with the exception of the following modifications. In this experiment, and unlike the first, if the ball reached the right side of the screen the trial continued until the subject released the response key. On trials where the ball moved off the screen, instead of displaying the ball’s final position as feedback, a large yellow arrow was displayed pointing to the right side of the screen. Subjects were instructed that this feedback indicated that the ball had moved off the edge of the display. The point value associated with moving past the edge of the display was the same as the rightmost region for the penalty configuration on each trial.

In addition, subjects did not complete any practice trials in the second experiment. Instead, the entire experiment consisted of attempting to maximize a point score in the presence of penalty and reward regions. This modification was made for two reasons. First, the absence of scores during the practice phase of the experiment made the performance objective for the task mathematically underspecified—subjects were told to aim for the center of the targets, but were given no instructions regarding the value or cost of missing the target by a small amount. To model this portion of the experiment it was necessary to hypothesize the form of the loss function adopted by subjects. The chosen loss function, a quadratic distance penalty, was consistent with other estimates of loss functions in motor control (Körding & Wolpert, 2004), but could not be empirically measured in the Bayes’ Ball paradigm. In contrast, the penalty and reward regions explicitly and fully define the performance objective for the subjects. Second, by eliminating the practice phase of the experiment, it was possible to collect more trial data from each subject within a 1-hour experimental session.

After each block of 225 trials, overall feedback was given to the subjects in the form of a high score list. This feedback consisted of a bar graph of the subject’s average score over the last block of 225 trials, compared to the mean score of each previous subject in the experiment for the corresponding block of 225 trials (the structure of the blocks is described in detail in the next section). For the first subject, his or her performance was compared to data from 5 previous pilot subjects (these included the experimenter’s scores). Later subjects saw scores from all previous subjects. If a subject’s mean score
was in the top half of scores across all participants, a message displayed “Keep up the
good work!”, otherwise the message read “Try to do better on the next block!”
A sample of this feedback window is shown in Figure 5-1. The purpose of the feedback
was to motivate subjects to perform their best in the experiment despite the repetitive
nature of the task.

Figure 5-1. Sample feedback given to participants after each block of 225 trials.

Finally, at the end of the experiment, subjects completed a short survey designed to
test their explicit knowledge of the optimal compensation strategy. The survey consisted
of a screenshot of each distance and penalty configuration. The mouse cursor controlled
the position of the ball, subject to the constraint that the ball only moved along the same
horizontal track as during the course of the experiment. In addition, the ball remained
visible even when subjects moved the ball past the start of the occluding window.
Subjects were instructed to place the ball (using the mouse) at the location they
estimated was the optimal position to aim for in order to maximize their score for the
current trial configuration. Subjects indicated their estimate of the optimal strategy by
clicking the mouse on a location on the screen. The nine combinations of distance and
penalty configuration were presented in random order for each subject.

5.2.3 Design

The experiment included one between-subject manipulation. For half of the subjects,
artificial noise was added to the ball’s final location on each trial, before the ball position
was revealed to the subject. This is referred to as the noise condition. The value of noise
on each trial was sampled from a zero-mean Gaussian distribution with standard
deviation equal to half the target width, a value that equaled 41 pixels. If subjects released the response key while the ball was still visible (before it reached the occluding window), no noise was added to its position\(^\text{15}\). In addition, if subjects released the key immediately after the ball had moved behind the occluding window, the sampled noise was constrained so that it would not cause the resulting ball position to appear to the left of the occluding window. The remaining half of subjects interacted with a version of the task where no additional noise was added to the ball position (referred to as the no-noise condition). Subjects were not informed of the possibility of artificial noise in the experiment, and no visual properties of the task could indicate to subjects the presence or absence of artificial noise. Indeed, the experimenter found it difficult to determine which version of the task he was interacting with based purely on the feedback given after each trial.

Like the first experiment, targets varied according to their distance (near, middle and far) as well as their penalty configuration (penalty-before, both-penalty, and penalty-after). However, in this experiment the trials were presented in nine blocks of 225 trials each, where all the trials in a block consisted of a single penalty configuration and target distance. Blocks were further aggregated into groups of three consecutive blocks (for convenience they will be referred to as super-blocks). For all subjects, each super-block used a constant target distance, but the three blocks composing a super-block varied in their penalty configuration. The first block in each super-block was constrained to use the both-penalty configuration. The second and third blocks in a super-block used the penalty-before and penalty-after configurations, presented in randomized order. Thus, blocks 1, 4, and 7 for all subjects were constrained to be both-penalty trials. The order of the three target distances for the three super-blocks was counterbalanced across subjects using a Latin square design. These within-subject manipulations are summarized in Table 5-1.

\(^{15}\) This was done to prevent the occurrence of a visible jump in the ball’s position.
Table 5-1. Summary of the within-subject manipulations for Bayes’ Ball experiment 2. The order of distances was counterbalanced across subjects, while the order of penalty configurations was randomized subject to the constraint that both-penalty trials occurred first in each sequence of three blocks.

<table>
<thead>
<tr>
<th>Super-block</th>
<th>Block</th>
<th>Distance (counterbalanced)</th>
<th>Configuration (randomized)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>(d_1 \in { \text{near, middle, far} })</td>
<td>(c_1 = \text{both-penalty})</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>(d_1)</td>
<td>(c_2 \in { \text{penalty-before, penalty-after} })</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>(d_1)</td>
<td>(c_3 \in { \text{penalty-before, penalty-after} })</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>(d_2 \in { \text{near, middle, far} })</td>
<td>(c_1 = \text{both-penalty})</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>(d_2)</td>
<td>(c_2 \in { \text{penalty-before, penalty-after} })</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>(d_2)</td>
<td>(c_3 \in { \text{penalty-before, penalty-after} })</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>(d_3 \in { \text{near, middle, far} })</td>
<td>(c_1 = \text{both-penalty})</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>(d_3)</td>
<td>(c_2 \in { \text{penalty-before, penalty-after} })</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>(d_3)</td>
<td>(c_3 \in { \text{penalty-before, penalty-after} })</td>
</tr>
</tbody>
</table>

The order of the target distances was counterbalanced across subjects to eliminate any confound between target distance and block number. The order of the penalty configurations was randomized rather than counterbalanced because counterbalancing penalty configuration as well as distance would have required an impractical number of between-subject groups.

5.2.4 Procedure

Subjects were tested individually. Each subject was given a brief (approximately 10 minute) slide presentation that explained the task and the operation of the response pad. Subjects were informed that the ball moved at the same speed on every trial. Subjects were instructed with the point value of each region, and were told to maximize their total point score. No strategy or advice was given to the subjects regarding how they could maximize their score. Subjects were informed that their scores would be recorded and added to a high score list at the end of the experiment. This was done to further incentivize subjects to perform to the best of their abilities.

Subjects began the experiment following completion of the slide presentation. Trials were self-paced, and given the velocity of the ball, each trial typically lasted less than
1200 ms. In total, each subject completed 1425 trials. The entire experiment was conducted in a single session and lasted for approximately 45 minutes.

5.3 Results

5.3.1 Mean points per trial

The possible score on each trial of the experiment was –200, 0, or +100 points. On average, across all conditions, subjects earned 10.18 points per trial (SE = 1.40). To investigate whether the score differed between conditions, a $2 \times 3 \times 3$ mixed design ANOVA was conducted, with noise condition (noise vs. no-noise), target distance (near, middle, or far), and penalty configuration (penalty-before, both-penalty, and penalty-after) as factors. The analysis revealed a significant three-way interaction between configuration, distance, and noise condition, $F(4, 184) = 5.269$, $MSE = 750.54$, $p < .001$. The mean scores for each group are plotted in Figure 5-2 and reported in Table 5-2.

A three-way interaction is characterized by a significant difference between two-way interactions among two or more cells in a design. In Figure 5-2, it can be seen that in the no-noise condition, the difference between penalty-after and penalty-before trials is significantly different for the middle target distance ($t(23) = 3.50, p < .05, \alpha = .025$ corrected for multiple comparisons), while the difference is not significant in the noise condition ($t(23) = 0.83, p > .05, \alpha = .025$).

The two-way interaction between penalty configuration and target distance was found to be significant ($F(4, 184) = 50.02$, $MSE = 7126.24$, $p < .001$). For near targets, the difference in scores between the both-penalty and penalty-after configurations was smaller than for the far targets ($t(47) = 11.56, p < .001, \alpha = .025$ corrected for multiple comparisons). This interaction might be expected, as regardless of distance, any reasonable decision strategy for the penalty-before and penalty-after trials is likely to yield a non-negative score (obtaining a zero score merely requires aiming as far as possible away from the penalty region). In contrast, for the both-penalty trials, the mean score would be expected to decrease with increasing target distance, since there is no longer a ‘safe’ strategy to avoid penalty regions.

The two-way interaction between target distance and noise condition was also found to be significant ($F(2, 92) = 13791.60$, $MSE = 13791.60$, $p < .001$). In the no-noise
condition, subjects achieved a higher score for the near targets than subjects in the noise condition (compare Figure 5-2, left versus right panel). However, the rate of decrease in score with increasing target distance was greater for the no-noise condition than for the noise condition. This effect is to be expected given the nature of the noise manipulation. Imagine, hypothetically, that in the noise condition the standard deviation of artificial noise added to the ball’s position was made arbitrarily large. If this were the case, then no matter where the subject aimed for on each trial, the resulting ball position would be effectively completely random. In this case, the expected score on each trial would not vary at all with increasing target distance. With a smaller noise magnitude, performance would be expected to improve for the closer targets, but the size of the effect would be moderated, resulting in the shallower slope observed in Figure 5-2 (right panel).

Figure 5-2. Mean scores for each combination of target distance, penalty configuration, and noise condition. (Left) The data for the no-noise condition. (Right) Data for the noise condition. Error bars indicate 95% confidence intervals.
Table 5-2. Mean scores for each combination of noise condition, penalty configuration, and target distance.

<table>
<thead>
<tr>
<th>Noise cond.</th>
<th>Config.</th>
<th>Distance</th>
<th>Mean score</th>
<th>+/- 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>No-noise</td>
<td>Both-penalty</td>
<td>Near</td>
<td>45.56</td>
<td>8.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Middle</td>
<td>-3.06</td>
<td>8.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Far</td>
<td>-64.78</td>
<td>8.46</td>
</tr>
<tr>
<td>Penalty-after</td>
<td>Near</td>
<td>71.37</td>
<td>5.75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>52.97</td>
<td>4.42</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Far</td>
<td>18.03</td>
<td>5.93</td>
<td></td>
</tr>
<tr>
<td>Penalty-before</td>
<td>Near</td>
<td>72.27</td>
<td>4.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>43.70</td>
<td>5.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Far</td>
<td>13.73</td>
<td>5.36</td>
<td></td>
</tr>
<tr>
<td>Noise</td>
<td>Both-penalty</td>
<td>Near</td>
<td>-33.33</td>
<td>8.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Middle</td>
<td>-51.67</td>
<td>8.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Far</td>
<td>-87.06</td>
<td>8.46</td>
</tr>
<tr>
<td>Penalty-after</td>
<td>Near</td>
<td>28.07</td>
<td>5.75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>20.53</td>
<td>4.42</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Far</td>
<td>6.57</td>
<td>5.93</td>
<td></td>
</tr>
<tr>
<td>Penalty-before</td>
<td>Near</td>
<td>31.03</td>
<td>4.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>17.87</td>
<td>5.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Far</td>
<td>1.47</td>
<td>5.36</td>
<td></td>
</tr>
</tbody>
</table>

In addition to the interactions reported above, the main effects of target distance, penalty configuration, and noise condition were also significant. For the effect of target distance ($F(2, 92) = 437.67, \text{MSE} = 108034.06, p < .001$), mean score significantly decreased for each increasing distance. For the effect of penalty configuration ($F(2, 92) = 1016.06, \text{MSE} = 196026.62, p < .001$), mean score was significantly lower for the both-penalty trials. Scores did not significantly differ between penalty-before and penalty-after trials, with the exception of the small but significant difference in the no-noise condition at the middle target distance (see Figure 5-2, left panel). For the effect of noise condition ($F(1, 46) = 157.29, \text{MSE} = 133403.62, p < .001$), subjects scored significantly worse in the artificial noise condition (mean score = –7.39, SE = 1.98) than in the no-noise condition (mean score = 27.75, SE = 1.98).

5.3.2 Evidence for score improvement across blocks

On blocks 1, 4, and 7 of the experiment, the penalty configuration was constrained to be both-penalty for all subjects (the target distance however, was counterbalanced). As a test for ongoing score improvements over the course of the experiment, a $2 \times 3$
repeated-measures ANOVA for mean score was conducted with noise condition and block number (1, 4, or 7) as factors. The analysis revealed no evidence for any change in mean score over the course of these three blocks (F(2,92) = .132, MSE = 293.82, p = 0.88). This lack of effect does not prove that no learning occurred in the experiment, but does suggest that any learning that occurred in the experiment was likely limited to the early blocks of the experiment.

5.3.3 Delta

In keeping with the first experiment, a dependent measure delta was defined as the difference between the ball’s stopping position and the center of the reward region, measured in mm: \( \Delta = x_{\text{ball}} - \text{target-center} \). For the noise condition, the un-corrupted ball position was used in the calculation of delta in order to avoid reducing the sensitivity of the analysis. According to this measure, positive values indicate stopping to the right of the center of the reward region, and negative values to the left. Mean delta values for each combination of experimental conditions are reported in Table 5-3 and plotted in Figure 5-3.

![Figure 5-3](image)

*Figure 5-3. Mean delta values for each condition. (Left) Data from the No-noise condition (Right) Data from the Noise condition. Error bars indicate 95% confidence intervals.*
Table 5-3. Mean delta values for each condition.

<table>
<thead>
<tr>
<th>Cond.</th>
<th>Config.</th>
<th>Distance</th>
<th>Delta (mm)</th>
<th>+/– 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>No-noise</td>
<td>Both-penalty</td>
<td>Near</td>
<td>1.12</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Middle</td>
<td>-1.74</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Far</td>
<td>-2.53</td>
<td>1.45</td>
</tr>
<tr>
<td>Penalty-after</td>
<td>Near</td>
<td>-2.41</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>-4.36</td>
<td>1.23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Far</td>
<td>-9.83</td>
<td>2.05</td>
<td></td>
</tr>
<tr>
<td>Penalty-before</td>
<td>Near</td>
<td>3.46</td>
<td>1.75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>3.10</td>
<td>1.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Far</td>
<td>6.71</td>
<td>2.82</td>
<td></td>
</tr>
<tr>
<td>Noise</td>
<td>Both-penalty</td>
<td>Near</td>
<td>2.66</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Middle</td>
<td>-0.14</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Far</td>
<td>-1.48</td>
<td>1.45</td>
</tr>
<tr>
<td>Penalty-after</td>
<td>Near</td>
<td>-3.97</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>-7.53</td>
<td>1.23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Far</td>
<td>-10.71</td>
<td>2.05</td>
<td></td>
</tr>
<tr>
<td>Penalty-before</td>
<td>Near</td>
<td>6.84</td>
<td>1.75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>5.31</td>
<td>1.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Far</td>
<td>11.21</td>
<td>2.82</td>
<td></td>
</tr>
</tbody>
</table>

A $2 \times 3 \times 3$ ANOVA of delta by noise condition, target distance, and penalty configuration indicated a significant interaction between penalty configuration and target distance ($F(4,184) = 46.89$, MSE = 6179.95, $p < .001$). As would be expected, delta values were relatively unchanged over increasing target distance for the both-penalty trials. For the penalty-before trials, delta was positive at near distances and increased with increasing distance, while for penalty-after trials, delta became increasingly negative with increasing target distance. Importantly, this effect differs from the first experiment, where delta decreased, rather than increased, for the penalty-before as well as both-penalty trials with increasing distance. The absence of this effect in the second experiment demonstrates that the overall bias observed in the first study can be eliminated by changing the experiment to use a blocked design. For the both-penalty trials, there is some evidence that subjects overshot the near target, but the size of this bias was only 1.05 mm, and given the 95% confidence intervals around the mean, was just barely interpretable as significantly different from zero.

A significant interaction between penalty configuration and noise condition was also observed ($F(2,92) = 8.79$, MSE = 3622.22, $p < .001$). In the noise condition, the different
penalty configurations had a larger effect on delta than in the no-noise condition. This demonstrates that subjects increased the magnitude of their compensation away from the penalty targets in the presence of artificial noise. Note that delta was computed using the uncorrupted ball position on each trial rather than the noise condition, and therefore differences in delta between the two conditions cannot be due to the physical manipulation itself, but rather must instead be attributed to the human subjects’ response to the manipulation.

The main effects of distance, configuration, and noise condition were also significant. In terms of distance (F(2,92) = 28.16, MSE = 3630.31, p < .001), mean delta was slightly positive for the near target distance (mean = 1.28 mm, SE = 1.11), and slightly negative for the middle (mean = -0.89 mm, SE = 0.90) and far (mean = -1.11 mm, SE = 1.48) targets. These are consistent with a centering bias, but as was previously noted, the magnitude of the bias is almost inconsequential. In particular, the mean delta was not significantly different between the middle and far targets. As far as the main effect of penalty configuration (F(2,92) = 199.22, MSE = 82044.73, p < .001), delta was positive for penalty-before trials, negative for penalty-after trials, and not significantly different from zero on both-penalty trials. For the main effect of noise condition (F(1,46) = 4.10, MSE = 1448.29, p = .049), mean delta was slightly lower in the no-noise condition (mean = -0.72 mm, SE = 1.28) than in the noise condition (mean = 0.24 mm, SE = 1.28), although the difference just barely reached significance.

5.3.4 Compensation

In the previous experiment, compensation was defined as the difference between ball position on a given trial and mean ball position on both-penalty trials. In addition, the sign was reversed for penalty-after trials so that positive values consistently indicated compensating away from the penalty region. This dependent measure was computed in order to subtract out the overall bias observed in the data, and concentrate on the effect of varying penalty configuration on stopping position. However, in this experiment little or no bias was observed; consequently subtracting out mean stopping position on both-penalty trials (which was essentially zero) would not add any additional insight. Therefore analysis of compensation is not reported for this experiment.
5.3.5 Transfer trials

One of the reasons for using the blocked design in this experiment was to compare performance at the end of the first block to performance at the beginning of the second block. Recall that block 1, for all subjects, consisted of both-penalty trials, while in block 2 the configuration was either penalty-before or penalty-after, chosen randomly. If subjects learned to perform the task by maximizing expected reward, without learning about their variability, then they would not be able to shift their aiming position on the first trial of the second block based on the penalty configuration.

Changes in delta from trials 216 to 235 (i.e., the last 10 trials of block 1 and the first 10 trials of block 2) were analyzed using a $2 \times 2 \times 20$ mixed-design ANOVA, with noise condition, transfer configuration (to-Penalty-before versus to-Penalty-after) and trial number (216 to 235) as factors. The analysis revealed a significant interaction between trial number and transfer condition ($F(19,817) = 3.35, MSE = 13615.23, p < .001$). The main effect of trial number was also significant ($F(19,817) = 4.65, MSE = 18890.67, p < .001$). No effects or interactions with noise condition were found to be significant (all $p > .05$). Two subsequent one-way ANOVAs on delta by transfer condition revealed that on the first trial of block 2, mean delta differed according to the transfer condition ($F(1,45) = 9.94, MSE = 37801.36, p < .05$, $\alpha = .025$, adjusted for multiple comparisons), but did not differ for the trial immediately preceding the switch ($p > .05$). On the first trial after transfer, delta was more negative for penalty-after trials than for penalty-before trials. These results are illustrated in Figure 5-4. For the sake of legibility, error bars are only displayed for the penalty-after transfer condition.
Figure 5-4. Mean delta by trial number and transfer condition. The transfer from block 1 to block 2 occurred after trial 225. Error bars on the penalty-after transfer condition indicate 95% confidence intervals.

One peculiar feature of the results is that on trial 226, subjects undershot the target regardless of penalty configuration. By the second trial of block 2, this bias had disappeared. The effect may be due to the intervening feedback window displayed at the end of block 1, or it may be due to the sudden change in penalty configurations on the first trial of block 2; without further information or replication, the precise origin of the effect cannot be accounted for. Despite this, even on the first trial of the second block, significant differences emerged between penalty-before and penalty-after configurations. By design of the experiment, this difference could not be attributed to learning based on reinforcement from previous trials.

5.3.6 Post-experiment survey results

At the end of the experiment, subjects were asked to indicate for each distance and penalty configuration the location they estimated was the optimal position to aim for in order to maximize their score for that configuration. Subjects indicated their estimates by clicking the mouse on a location on the screen. A dependent measure referred to as survey-delta was computed as the difference between the subjects’ response and the target center for each penalty and configuration—this measure is equivalent to the delta measure computed for trials during the experiment. The nine combinations of distance and penalty configuration were presented in random order for each subject. Figure 5-5 compares mean delta empirically observed in the experiment (top row) to the results of
the post-experiment survey (bottom row). The graphs in the left column are subjects in the no-noise condition, while the graphs in the right column are from the noise condition.

![Graphs showing comparison of mean delta empirically observed during the experiment (top row) versus subjects’ estimates of the optimal strategy for the task (bottom row). The left column is data from the no-noise condition, while the right column is data from the noise condition. Error bars indicate 95% confidence intervals.](image)

Figure 5-5. Comparison of mean delta empirically observed during the experiment (top row) versus subjects’ estimates of the optimal strategy for the task (bottom row). The left column is data from the no-noise condition, while the right column is data from the noise condition. Error bars indicate 95% confidence intervals.

A $2 \times 3 \times 3$ mixed-design ANOVA of survey delta with noise condition, target distance, and penalty configuration as factors revealed a main effect of penalty configuration ($F(2, 92) = 80.05$, $MSE = 100703.88$, $p < .001$). This demonstrates that subjects understood the purpose of the survey questions; they knew that aiming for the center of the target was an optimal strategy for both-penalty trials, and aiming away from the penalty zone was a good strategy for the penalty-before and penalty-after trials. Critically however, the interaction between penalty configuration and distance was not
significant ($F(4,184) = 0.776$, $MSE = 61.81$, $p = .54$). Subjects were not explicitly aware that they needed to shift their aiming position to the right with increasing target distance for penalty-before trials, and shift to the left for penalty-after trials. The main effect of distance was similarly nonsignificant. In addition, no effects or interactions with noise condition were found to be significant (all $p > .05$), including the expected interaction between noise condition and penalty configuration.

5.3.7 Summary and discussion of results

In summary, this experiment extended the results of the first Bayes’ Ball experiment in five important ways. First, it served as a replication of the findings from the first experiment. Second, it introduced a blocked trial structure in order to test the hypothesis that a centering bias observed in the first experiment was due to sequential interference stemming from a randomized design. Based on the results of the second experiment, this bias was essentially eliminated from human performance, in support of the hypothesis.

Third, this experiment provided a strong test of whether subjects were in fact adapting to variability per se, or some other property of the task. This was accomplished by covertly introducing additional variability in the ball’s position for half of the subjects. Subjects in this noise condition increased their compensation away from penalty regions compared to subjects in the no-noise condition, confirming the hypothesis.

Fourth, the blocked design allowed for a strong test of whether subjects learned internal models of their variability, or whether they merely adjusted their aiming position based purely on trial-to-trial feedback. This was possible by constraining all subjects to complete the both-penalty configuration in the first block, and then examining compensation away from the penalty regions at the beginning of the second block. The results demonstrated that on the very first trial of the second block, subjects altered their aiming position as a function of the penalty configuration, ruling out the possibility that they learned without acquiring at least an approximate internal model of their variability.

Fifth, and finally, the results of the post-experiment survey revealed that subjects had no explicit awareness of the dynamics of the Weber law. Their responses in the survey indicate that they believed their ability to estimate the ball’s position was
invariant with respect to target distance. Similarly, subjects exhibited no explicit awareness of the noise manipulation: their estimated optimal aiming point did not vary depending on which noise condition they were in. Despite this lack of explicit awareness, in their actual performance of the task, subjects exhibited acute sensitivity to both the Weber law of interval timing, as well as the additional variability imposed by the noise condition. These findings suggest that the acquisition or exploitation of predictive models of embodied dynamics does not depend on an explicit intention or strategy to do so, but rather reflects the normal mode of human routine interactive behavior.

5.4 Ideal performer analysis

The derivation for the ideal performer analysis for the Bayes’ Ball paradigm was provided in the previous chapter and in Appendix B. The same procedure was used to obtain optimal decision strategies for each participant in the second experiment, with the exception of the following modifications. First, recall that the task environment was modified for the second experiment so that if the ball moved beyond the right edge of the screen the trial continued until the subject released the key, and the resulting score was equal to that of the rightmost point region. For the ideal performer analysis, this simply amounted to treating the rightmost point region on each trial as if it had infinite extent. The artificial noise introduced for half of the subjects did not change the procedure of the analysis, as the variability parameters $\sigma_0$ and $\sigma_1$ were separately fit for each subject. From the model’s perspective as well as the humans’, any variability in the ball’s position cannot be definitively attributed to internal or external origins.

The output of the ideal performer analysis consisted of a prediction, for each subject, of the optimal aiming position for each combination of target distances and penalty configurations. In addition, the ideal performer analysis predicted the optimal scores that the human subjects could theoretically achieve given their level of estimation uncertainty.
Figure 5-6. Comparison of human subject and ideal performer delta, for each combination of distance, penalty configuration, and noise condition. Human data are presented in the top row, model predictions are in the bottom row. The left column shows data from the no-noise condition, while the data in the right column is from the noise condition.

Figure 5-6 compares the empirical results (top row) to the predictions of the ideal performer analysis (bottom row) in terms of mean delta. The left column is data from the no-noise condition, while the right column is from the noise condition. For the no-noise condition, the ideal performer predictions are almost identical to those of the first experiment, as would be expected given that the ideal performer model is insensitive to the order of trials. For the noise condition, the ideal performer model increased the magnitude of its compensation away from the penalty regions, as is also to be expected. The magnitude of the predicted optimal compensation was larger for the penalty-before trials than for the penalty-after trials. Unlike the first experiment, there was no penalty
associated with moving past the edge of the display. In addition, uncertainty would be expected to be greater when shifting to the right of a penalty than shifting to the left of a penalty region, given that the former requires estimating a larger time interval. These two facts account for the difference in compensation between penalty-before and penalty-after trials.

Across both conditions, the ideal performer analysis demonstrates that human subjects were slightly suboptimal in terms of the magnitude of their compensation. Yet despite the differences between observed and optimal compensation, there was virtually no difference between the observed and predicted optimal scores. On average, human subjects earned 10.18 points per trial. The ideal performer analysis predicted that the expected maximum obtainable score was 11.58 points per trial, a difference of less than two points per trial. Figure 5-7 (left panel) shows a scatterplot comparing predicted optimal and empirical scores. Each marker point in the graph indicates the mean score for a single subject at a particular combination of target distance and penalty configurations. The resulting correlation between the scores predicted by the ideal performer and human subjects is high ($r^2 = .91$).

![Figure 5-7](image)

**Figure 5-7.** (Left) Correlation between predicted optimal and empirical scores. (Right) Correlation between model-fit standard deviation in ball position and empirically observed standard deviation. The diagonal dashed lines indicate the unit slope corresponding to perfect correlation, the solid line gives the best fit linear regression.
The results obtained from the ideal performer analysis in terms of compensation and resulting score seem to imply a contradiction: subjects were below the ideal performer model in terms of the magnitude of their compensation away from penalty regions, but yet performed at optimal levels in terms of their resulting scores. One possibility is that the ideal performer did not accurately capture the constraints on human performance. That is, subjects could match performance of the ideal performer model and exhibit less compensation if the ideal performer analysis overestimated each subjects’ variability. The right panel of Figure 5-7 shows the correlation between the model’s predicted standard deviation in ball position for each subject, distance, and penalty combination, and the empirically observed standard deviation in ball position. The correlation between predicted and observed standard deviation is also high ($r^2 = .83$). The linear regression indicates that the ideal performer analysis overestimated standard deviation for human subjects by a small amount, and this could potentially account for some of the apparent contradiction. Reducing the variability for the ideal performer models would have the effect of boosting their expected scores in the task, but would also reduce the magnitude of their compensation, bringing the model closer in line with the empirically observed compensation.

Another possible explanation has nothing to do with the ideal performer analysis, but rather has to do with properties of the Bayes’ Ball paradigm itself. For both-penalty trials, there is a narrow range of aiming positions that result in optimal or near-optimal expected scores. For the penalty-before and penalty-after configurations, there is a much wider range of aiming positions that will result in a near-optimal score, and this effect is magnified for the farther target distances. Figure 5-8 shows the computed expected utility curves for a single subject on the penalty-before trials at each of the three target distances. The curves were obtained based on the ideal performer model fit to this subject, and indicate the expected score that the subject would receive as a function of his or her decision strategy (mean aiming position). As can be seen in the graph, for each target distance there is a peak corresponding to the maximum expected utility. Aiming to the left of this peak results in negative expected scores (since they are penalty-before trials). Aiming to the right also results in a lower score, but the falloff in expected utility is much more gradual. Further, with increasing target distance, the peak of the utility
curve becomes increasingly flat. At the farthest target distance and for this particular subject, there is a range approximately 200 pixels wide (52.68 mm), inside of which any decision strategy the subject adopted would result in a mean score that differed from the optimal by less than 2 points.

Figure 5-8. Expected utility curves for a single subject on the penalty-before configuration at each of the three target distances (distance measured in pixels).

This finding, combined with the fact that human subjects matched the predicted optimal scores in the task, suggests that human subjects are not suboptimal in the Bayes’ Ball paradigm, but rather they are optimal to within the sensitivity of the paradigm itself as a probe of their performance.

5.5 Cognitive model based on the adaptive internal model framework

The second Bayes’ Ball experiment was motivated, in part, based on predictions from the cognitive model developed to account for that data. In particular, the biases observed in the first experiment were hypothesized to be due to two sources: a tendency to overshoot near targets and undershoot far targets, and a bias towards the center of each target. These biases were accounted for by introducing two parameters into the model. By changing the task from a randomized to a blocked design in the second experiment, the model makes quantitative predictions for the resulting performance. In particular, in a blocked design, the mean location of recent targets corresponds to the location of the current target. The parameter controlling the undershoot-overshoot bias was set to zero, while the parameter controlling the target center bias was left unchanged, as no
experimental manipulations were done to eliminate this bias. All other aspects of the model are identical to those described for the first experiment.

The model was then re-run, using the same experimental design as was used for human subjects. This included presenting distances in blocks, counterbalancing the three possible orders of target distances, and randomizing the order of penalty-before and penalty-after trials. In addition, simulations were run separately for the noise and no-noise conditions. Fifteen simulated subjects were run in each combination of target distance order and noise condition, resulting in 90 \((15 \times 3 \times 2)\) total simulated subjects. For the sake of computational tractability, simulations were only run through the end of the first target distance \((675\) trials per simulated subject)\(^{16}\). Predictions for the remaining two target distances were obtained from the agents that were trained using a different target distance order.

\(^{16}\) The shortened experiment required approximately 12 hours of computation time to complete. The number of simulated subjects \((15)\) was also limited for the sake of computational tractability.
Figure 5-9. Comparison of empirical (top row) and model predictions (bottom row) in terms of delta. The left column is data from the no-noise condition, while the right column is from the noise condition. Error bars on the human data are 95% confidence intervals.

Figure 5-9 compares the human and model data for the primary dependent measure of delta (recall that delta is the distance from the mean stopping position to the target center). The first 25 trials of the model data are excluded, as for this comparison asymptotic model performance is the primary focus, and it was found that the model’s performance had stabilized within approximately this amount of experience. No parameters were adjusted to fit the data, and the resulting model predictions closely match human performance in the experiment. For human and model subjects, the addition of artificial variability in the noise conditions induced small, but significant increases in the magnitude of compensation away from the penalty regions. This effect is
most pronounced for the near targets, where the magnitude of artificial noise was greatest compared to the intrinsic variability in time estimation.

5.6 Summary

This chapter presented the results of a second experiment designed to investigate the human ability to acquire and exploit internal models of the variability associated with estimating short time intervals. The results of the experiment provide further evidence for the main hypotheses of this dissertation—subjects were able to predict their individual levels of variability in the task, and exploit this knowledge to improve their task performance. In addition, the experiment demonstrated that subjects are able to incorporate predictions of not only their intrinsic variability in time estimation, but also extrinsic variability imposed in the form of artificial noise in the ball’s stopping position.

The blocked design of the experiment enabled a strong test of whether subjects were acquiring internal models of their variability, or instead learning a mapping from time estimates to expected reward (as would be predicted by a reinforcement learning model). For all subjects, trial 226 was the first trial on which asymmetrical penalties required shifting away from the center of the reward region. Up until this point, all subjects were reinforced only for aiming for the center of the target. It was observed that on trial 226, subjects shifted their mean aiming position in a manner adaptive to the penalty configuration (delta was greater for penalty-before trials than for penalty-after trials). The cognitive models based on the adaptive internal model framework were similarly immediately responsive to changes in trial stimuli.

A post-experiment questionnaire demonstrated that subjects’ implicit performance in the task was more optimal than their explicit judgments of the optimal task completion strategy. When explicitly asked for the optimal strategy, subjects showed no awareness of the Weber law of interval timing, or of the artificial noise manipulation.

An ideal performer analysis was conducted for the second experiment, and demonstrated that human performance in the experiment was optimal, subject to the constraints of time estimation variability imposed by the Weber law. Optimal performance was obtained despite subjects demonstrating less compensation away from penalty regions than predicted by the ideal performer analysis. Subsequent investigation
showed that this finding is likely the result of an inherent limitation of the Bayes’ Ball paradigm. As variability increases (at the farther target distances), the peak of the expected value curve grows increasingly flat. Since the goal for subjects was to maximize their score, any strategy that achieved this is functionally equivalent.

A cognitive model, based on the adaptive internal model framework, was also used to generate quantitative predictions for the second experiment. The cognitive model differed from the ideal performer analysis in that they were required to learn to perform the task on the basis of experience in a cognitively plausible manner, whereas the ideal performer analysis assumed both omniscience regarding its dynamics, and infinite mathematical competence. The cognitive model was identical to the one from the first experiment, with the exception of eliminating one of the two parameters. The justification for this change was a modification to the paradigm used in the second experiment. With no free parameters, the model was able to provide an excellent fit to human data, providing further evidence that the mechanisms underlying the model’s performance (the acquisition of internal models of embodied dynamics) are consistent with human performance in the task.

The Bayes’ Ball paradigm is similar in many respects to previous studies demonstrating acute awareness of the noise inherent to the motor system in rapid pointing tasks (Maloney et al., 2007; Trommershäuser et al., 2006). The principle difference is that the Bayes’ Ball paradigm was used here to investigate a cognitive, rather than motor source of uncertainty. Instead of losing points due to motor errors, subjects in Bayes’ Ball lost points due to uncertainty in the amount of elapsed time on each trial. Recently, a similar study has been conducted to investigate optimality in controlling the temporal duration of a physical movement (Hudson, Maloney, & Landy, 2008). In this experiment, subjects were required to reach and hit a target with varying temporal duration of movement. Movements that were too short in duration versus too long incurred asymmetrical penalties and rewards. The investigators found that in this experiment subject performance was nearly optimal, in keeping with the findings from the Bayes’ Ball studies reported here. Interestingly, there is evidence that at least one subject in their experiment demonstrated a consistent bias towards the center of the
reward interval (see Hudson et al., 2008, Figure 4). Expected gain curves for the remaining four subjects are not reported in their work.

On the one hand, the study by Hudson et al and the two experiments reported in this dissertation provide converging evidence for the same capacity to incorporate knowledge of temporal uncertainty in behavior. On the other hand, the study by Hudson et al is overly narrow in its interpretation of this ability. They argue that this capacity is a feature of the motor planning system, rather than a general capacity of an adaptive cognitive system. In contrast, the Bayes’ Ball studies have reduced the role of the motor system in task performance to controlling the release of a single key press. Hence, this work extends the interpretation of prior work from a narrow focus on the optimality of the motor system to a consideration of internal models of embodied dynamics as a general capacity of the human cognitive system in routine interactive tasks.
6. Experiment 3: ETA

6.1 Overview

Experiments 1 and 2 in this dissertation demonstrated the human capacity to acquire and exploit probabilistic internal models of the dynamics of ongoing cognitive, perceptual, and motor activity. However, the Bayes’ Ball paradigm itself was rather simple compared to the complexity and variety of activities humans routinely engage in on a routine basis. In particular, the task environment severely constrained the type of behavior that subjects could exhibit, by limiting the space of possible interaction to pressing and releasing a single button on each trial. By constraining the task environment so severely, the only remaining control dimension (the timing of the key release) might have been rendered especially salient or relevant to the experimental participants. That is, subjects may have realized that controlling the timing of the keypress may be an important component of task performance by virtue of it being the only aspect of the task over which they had any control. In most natural tasks, the environment does not restrict the space of control dimensions to just the relevant dimensions necessary for optimal performance. In addition, the system of point values assigned to various trial outcomes may have served to make the task somewhat unnatural to subjects. In most tasks outside of video games, maximizing arbitrary ‘points’ is not a useful objective.

Thus, the purpose of this experiment was to generalize the previous results to a more complex task environment that imposed fewer constraints on the space of available strategies, and adopted a more naturalistic performance objective (complete the task quickly and accurately). The ETA paradigm (short for estimated time of arrival) required subjects to coordinate a reaching movement with a cognitive decision making process that provided the eventual reaching target for the motor system. The difficulty of the cognitive decision task was manipulated by the use of a task switching paradigm, and the key question is whether and how the motor system would adapt to the properties of the concurrent cognitive task.

Previous research on task switching has shown that when humans frequently alternate between two or more activities, cognitive performance on each of the activities
can be slowed, a finding that is referred to as switch cost. Although the basic phenomenon itself has proven highly stable, there is wide disagreement over the advantages and disadvantages of various paradigms for measuring switch cost (see Altmann, 2007; Gopher, Armony, & Greenshpan, 2000; Logan & Bundesen, 2003; Pashler, 2000; Rogers & Monsell, 1995; Sims, 2003). The task switching paradigm adopted for this research is known as the explicit task-cueing procedure (Koch, 2001). In the version of the paradigm used here, pairs of colored objects are presented to the subject, and the task is to judge whether the two objects are the same or different according to a dimension specified by the task cue. The three possible task cues used in this experiment are color, shape, and texture. For part of the experiment, the trials were grouped into blocks where each trial in a block used the same task cue (eliminating task switching cost), whereas other blocks consisted of a mixture of all three task types. By using three possible tasks, it is possible to present blocks of switch trials where the task changes on every trial without making the task for the next trial predictable.

Responses in the experiment are made by moving the mouse cursor from a starting position, through a waypoint, to one of two response circles. On all trials, the initial movement from the start circle to the waypoint is the same, but the appropriate response circle is determined by the outcome of the object discrimination task. An additional restriction is that subjects must move from the waypoint to the response circle within a short timeout period or else the trial is counted as incorrect. The principal dependent measures in the experiment concern whether the subjects’ motor behavior on the initial segment of the movement demonstrates sensitivity and adaptation to the dynamics of the cognitive discrimination task. In particular, consider three possible strategies for completing each trial:

*Strategy 1—*independent serial strategy. After seeing the task cue and stimuli, the subject first decides the correct answer, and only then begins moving the mouse from the start circle to the waypoint. This strategy is suboptimal from the perspective of completing each trial as fast as possible, since it results in idle time for the motor system while the cognitive system is determining the correct answer. Presumably, the subject could have begun moving the mouse before or while deciding whether the two objects
are the same or difference, since the initial movement to the waypoint is the same regardless of the final answer.

*Strategy 2—independent parallel strategy.* The subject begins moving the mouse as soon as the two objects are displayed, and before he or she knows the correct answer for the current trial. This strategy is more time efficient than an independent serial strategy, since the initial reaching movement (from the start circle to the waypoint) is the same regardless of the answer, and it eliminates idle time in the motor system. However, the danger is that the mouse cursor will reach the waypoint before the subject knows the answer. In this case, the timeout period is initiated and the subject may not have time to determine the answer and reach the appropriate response circle before the timer runs out.

*Strategy 3—coordinated parallel strategy.* This strategy requires that the subject strategically adapt and coordinate the initial mouse movement such that the mouse cursor reaches the waypoint just after the subject determines the correct answer to the trial. This adaptation can take the form of strategic control over the initial movement latency (waiting a short interval before initiating the movement) or by varying the velocity or trajectory of the movement (moving the hand slower to allow the cognitive system more time to determine the answer). Importantly, this strategy requires that the subject be able to predict the temporal dynamics of both the cognitive discrimination task and the perceptual-motor reaching task. This strategy is assumed to be optimal for the experiment, since it minimizes both idle time in the motor system, as well as the possibility of failing to register a response within the response interval.

### 6.2 Hypotheses

The predictions for this experiment follow from the main hypotheses of this dissertation. Namely, it is predicted that subjects will acquire and exploit internal predictive models of the temporal dynamics of the cognitive discrimination task, as well as the motor dynamics associated with controlling the mouse in the experiment. The use of internal predictive models in the ETA paradigm would be manifest in evidence that subjects timed their initial movement from the start circle to the waypoint to coordinate with the anticipated temporal duration of the cognitive decision-making process. In particular, movements should be initiated later (higher latency), or executed more slowly (lower
velocity) in mixed blocks of trials compared to pure blocks. These differences should emerge in each trial before the cognitive decision task is completed. Further, the effects of task-switching on motor performance should disappear for subjects in a control condition where a parallel strategy is no longer optimal (described in the Method section, below). Trial completion time in the control group should be slower overall, due to the fact that subjects adopt a serial rather than parallel decision strategy.

6.3 Method

6.3.1 Participants

Fifty-six undergraduate students at Rensselaer Polytechnic Institute volunteered to participate in the experiment in exchange for course credit. The subjects’ mean age was 19.1 years (SD = 1.2). Eleven of the subjects were female. All subjects used their right hand to control a computer mouse.

6.3.2 Apparatus and Materials

The experiment was conducted on an Apple G4 computer running version 10.4 of the Macintosh operating system. All experimental stimuli were displayed on a 17” LCD monitor set to 1280×1024 screen resolution. Participants interacted with the experimental software using a standard Apple 1-button USB optical mouse (model number M5769). The sampling rate for this mouse was 100 Hz. The experimental software was implemented in the lisp programming language using LispWorks® version 4.4.

A diagram of the experimental interface is shown in Figure 6-1. At the start of each trial of the experiment, a small circle was displayed (diameter = 7.90 mm), referred to as the start circle. Subjects initiated each trial by moving the mouse cursor inside the start circle. At this point, the mouse cursor was hidden, and a task cue was displayed, located 26.34 mm above the start circle. The task cue consisted of two letters: “CO”, “SH”, or “TX”. These stood for color, shape, or texture, respectively. The task cue indicated to the subject which task was to be performed on the current trial. After waiting for a cue-stimulus-interval (CSI) lasting for 500 ms, the trial stimuli were presented at a location just below the task cue, and above the start circle. The trial stimuli consisted of two
simple geometric objects. The set of possible objects varied along three dimensions: their shape (circle or square), color (red or blue), and texture (solid or striped). Two objects from this set were displayed side by side, separated by 26.34 mm (center-to-center distance). Each object was 13.17 mm in diameter. The task cue remained visible after the trial stimuli were presented. As soon as the trial stimuli were displayed, the mouse cursor was relocated to the center of the start circle and made visible.

The experiment also displayed either one or two (depending on task condition) waypoint circles, and two response circles (diameter = 15.81 mm). Above one of the response circles the word “SAME” was displayed, and above the other the word “DIFFERENT” was displayed. The waypoint and response circles remained visible throughout each trial. At the start of each trial, the response circles were shaded a solid gray color, while the waypoint circles were solid white. Once the subject moved the mouse cursor from the start circle into a waypoint circle, the response circles changed color (from gray to solid white). A response timeout of 500 ms also began once the mouse entered the waypoint. The subject then moved the mouse cursor from the waypoint into the appropriate response circle, depending on the task cue and trial stimuli. If the subject failed to enter the waypoint circle (they either forgot or missed), the response circles remained shaded in gray and did not register the subject’s response. If the task cue for a given trial was “SH” (which stood for shape), and the two objects were both squares, then the correct response was “SAME”. After entering the waypoint, subjects would then have to move the mouse to this response circle within the 500 ms timeout period, or else the trial was ended and counted as incorrect. At the end of each trial, feedback was given to the subject. If they had registered the correct response, their total trial time was displayed in ms, and a cheerful tone was played (via headphones). If they had registered an incorrect response or failed to respond within the response timeout, the word “INCORRECT” was displayed and a buzzer sound was played. Figure 6-1a shows the sequence of events for each trial of the experiment. Figure 6-1b and c show the layout of the task interface for the two experimental conditions.
Figure 6.1. (a) Sequence of events on each trial. Subject-triggered events are shaded in gray. (b) Layout of the task interface for the bilateral experiment condition, after the mouse entered the start circle. (c) Task interface for the unilateral experiment condition.

The experimental software time-stamped and recorded each subject-generated event (such as the mouse entering and leaving the start circle, etc.). In addition, the experimental software recorded the location of the mouse cursor every 10 ms, starting from the moment the cursor left the start circle, and ending when the mouse cursor reached one of the two response circles.

6.3.3 Design

Each subject completed 660 trials in the experiment, grouped into two blocks of 330 trials. The experiment included one between-subject manipulation, the response mode for the interface. In one condition (the bilateral condition), the two response circles were on opposite sides of the start circle, and there was a waypoint on either side of the start circle (Figure 6-1b). In this condition, on some trials subjects were required to move the mouse to the left, while other trials required moving the mouse to the right. For the other condition (the unilateral condition), both response circles were to the right of the start
circle, and there was only a single waypoint (Figure 6-1c). In this condition, regardless of the task cue or stimuli, the required movement from the start circle to the waypoint circle was identical. The distance between the start and waypoint circles, and between the waypoint and response circles was held constant across the two conditions. An illustration of the two response conditions is provided in Figure 6-1b and c.

In addition to the between subject manipulation, the experiment also included two within-subject factors, the trial task, and the context. On each trial the task was either color, shape, or texture discrimination. In the pure context, trials were grouped into runs containing only a single task type. For example, subjects completed 110 shape discrimination trials, followed by 110 color discrimination trials, followed by 110 texture discrimination trials. The order of the three groups of tasks was randomized. In the mixed context, the block contained 110 trials of each of the three tasks, but the order of the tasks were randomized, subject to the constraint that the same task was never used for two trials in a row. For example, a valid sequence of tasks in the mixed block is color, shape, color, texture, etc. Each subject completed both a mixed block and a pure block, and the order of the two blocks was counterbalanced across subjects. The task stimuli were generated randomly such that the correct response on a given trial was equally likely to be “SAME” or “DIFFERENT”.

In summary, each subject completed 220 trials for each of three task types, half of which occurred in a mixed context, while the other half occurred in a pure context. For half the subjects (n = 28) the interface used a unilateral response mode, while the other half of the subjects used a bilateral response mode.

6.3.4 Procedure

Subjects were tested individually. Each subject was given a brief slide presentation that explained the task. Separate instructions were given for the bilateral and unilateral conditions. The instructions emphasized that the goal for the subjects was to complete each trial as fast as possible without making a mistake. No specific strategy for how this was to be accomplished was given. After starting the experiment, the experimenter remained in the room for the first 10 trials to ensure that subjects understood the task, but did not provide any feedback or guidance to subjects other than to re-explain the
mechanics of the task if necessary. During the remainder of the experiment the subjects wore headphones, through which the auditory feedback at the end of each trial was played.

### 6.4 Results

Subject performance was analyzed using a series of $2 \times 2 \times 3$ mixed-design ANOVAs, with response mode (bilateral vs. unilateral) as a between subject factor and task context (mixed vs. pure) and task type (color vs. shape vs. texture) as within-subject factors. Ten principal dependent measures were investigated:

- Overall percent correct
- Percent of trials resulting in timeout errors
- Percent of trials where an incorrect response was made
- Initial movement latency—the time between onset of the task stimuli and the time that the mouse cursor left the start circle
- Percent of movements made in the wrong direction (away from the correct waypoint)
- Movement speed—the ratio of mouse pathlength and movement duration
- Movement pathlength
- Movement duration from the waypoint to the response circle
- Velocity profiles—the time course of velocity over the movement
- Vertical deflection—deviation of the mouse from a straight horizontal trajectory between the start and waypoint circles

Properties of the mouse trajectories were extracted and analyzed as follows. The experimental software sampled the mouse position every 10 ms, beginning at the time the mouse cursor left the start circle, and ending the moment the cursor first entered a waypoint circle. The raw $x$ and $y$ trajectories were separately smoothed using cubic spline interpolation to eliminate aliasing artifacts, via the `smooth.spline` function in the R
statistical package\textsuperscript{17}. Time derivatives were taken of the smoothed $x$ and $y$ trajectories, and a tangential velocity profile was constructed using the formula

$$v(t) = \sqrt{x'(t)^2 + y'(t)^2}.$$  

The velocity profile was then resampled at 5% quantiles of the total movement duration. Mouse pathlength was computed by summing the incremental Euclidean distance traveled by the mouse on each sample of the raw data.

6.4.1 Overall percent correct

On average, subjects completed 87\% of trials correctly. The proportion correct for each response mode, task and context are reported in Table 6-1. A two-way interaction between context and response mode was found to be significant ($F(1,54) = 5.641$, MSE $= 2.46E-02$, $p < .05$). This interaction is plotted in Figure 6-2. In terms of percent correct, there was no difference between the mixed and pure contexts in the bilateral response mode, but in the unilateral condition performance was significantly worse during the mixed block of trials.

<table>
<thead>
<tr>
<th>Response mode</th>
<th>Task</th>
<th>Context</th>
<th>Mean</th>
<th>+/- 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilateral</td>
<td>Color</td>
<td>Mixed</td>
<td>0.89</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>0.88</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Shape</td>
<td>Mixed</td>
<td>0.89</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>0.89</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Texture</td>
<td>Mixed</td>
<td>0.88</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>0.88</td>
<td>0.03</td>
</tr>
<tr>
<td>Unilateral</td>
<td>Color</td>
<td>Mixed</td>
<td>0.85</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>0.88</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Shape</td>
<td>Mixed</td>
<td>0.86</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>0.88</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Texture</td>
<td>Mixed</td>
<td>0.82</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>0.86</td>
<td>0.03</td>
</tr>
</tbody>
</table>

\textsuperscript{17} Cubic spline interpolation is frequently used for fitting motor trajectories, as the result of fitting is a polynomial equation. The temporal derivatives of the movement (speed and acceleration) can therefore be analytically determined with little effort from the resulting spline equations.
In addition to the interaction between context and response mode, the main effect of task type was found to be significant ($F(2,108) = 5.753$, $MSE = 8.62E-03$, $p < .05$). Subjects performed best on the shape discrimination trials (reaching 87.7% correct), followed by the color discrimination trials (87.3%), and the texture discrimination trials (86.0%). No interactions between task type and context or response mode were found to be significant. Although subjects on average performed significantly worse on texture discrimination trials, the magnitude of this difference was small (a 1.7% higher error rate).

### 6.4.2 Timeout errors

The preceding analyses indicated small but significant differences in overall percent correct that depended on the task type, as well as the interaction between context and response mode. Errors in this performance measure could have been due to two sources: choosing the wrong answer, or failing to respond within the timeout period. This section examines the percent of trials resulting in a timeout error. Overall, timeout errors occurred on 10.5% of trials, and thus accounted for 80.7% of the total errors recorded. However, no significant differences were found in the percent of timeout errors for either the task type, context or response mode (all $p > .05$). This suggests that the
differences in error rate between conditions were due to subjects registering incorrect responses, rather than failing to respond in time. This is investigated in the next analysis.

### 6.4.3 Incorrect response errors

Incorrect responses (successfully responding within the timeout period, but with the wrong answer) occurred on 2.5% of all trials, and accounted for 19.3% of the total errors recorded. The incorrect response rate for each combination of task conditions is reported in Table 6-2.

<table>
<thead>
<tr>
<th>Response mode</th>
<th>Task</th>
<th>Context</th>
<th>Mean</th>
<th>+/- 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilateral</td>
<td>Color</td>
<td>Mixed</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Shape</td>
<td>Mixed</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Texture</td>
<td>Mixed</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Unilateral</td>
<td>Color</td>
<td>Mixed</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Shape</td>
<td>Mixed</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Texture</td>
<td>Mixed</td>
<td>0.07</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>0.03</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Despite the low occurrence rate of these errors, an analysis of variance found the three way interaction between task, context, and response mode to be significant \((F(2,108) = 5.096, \text{MSE} = 1.265\text{E-03}, p < .05)\). The three-way interaction stems from a two-way interaction between task and context that was significant for the unilateral response mode \((F(2,54) = 9.06, \text{MSE} = 3.52\text{E-3}, p < .001)\), but nonsignificant for the bilateral response mode \((p > .05)\). In the unilateral condition, mixed trials resulted in a higher error rate than pure trials \((t(27) = 3.841, p < .05, \text{paired samples } t\text{-test})\), but the magnitude of the difference was largest for the texture discrimination trials. This interaction was unexpected, but further analysis of the data suggests that it is likely due to a single subject in the unilateral condition who reached a 31.8% incorrect response rate for the mixed block, texture discrimination trials. The second highest error rate across all conditions and subjects was roughly half this value, at 17.2%.
6.4.4 Data exclusion criteria

For the analyses that follow, trials resulting in timeout errors or incorrect response errors were excluded from the data. This was done to examine human performance on trials where inattention or lack of motivation were least likely to be influencing factors. In addition, trials on which the subject missed the waypoint were detected and excluded based on the recorded mouse trajectories. These were detected by flagging trials on which the maximum $x$ coordinate of the mouse cursor during the movement from the start circle to the waypoint exceeded the $x$ coordinate of the waypoint\textsuperscript{18}. Such trials accounted for 5.5% of the data. These data were excluded from analysis as performance on these trials would be expected to exhibit anomalous trial times, movements and velocity profiles.

6.4.5 Initial movement latency

Initial movement latency was defined as the time from when the trial stimuli were displayed and the mouse cursor was unhidden until the time that the subject moved the mouse cursor outside of the start circle. The mean latency for each experimental condition is reported in Table 6-3. 

\textsuperscript{18} Conversely, for trials where the waypoint was to the left of the start circle, the minimum $x$ coordinates were used.
Table 6-3. Mean initial movement latency by experimental condition.

<table>
<thead>
<tr>
<th>Response mode</th>
<th>Task</th>
<th>Context</th>
<th>Mean</th>
<th>+/- 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilateral</td>
<td>Color</td>
<td>Mixed</td>
<td>688.52</td>
<td>68.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>564.33</td>
<td>43.71</td>
</tr>
<tr>
<td></td>
<td>Shape</td>
<td>Mixed</td>
<td>701.44</td>
<td>74.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>561.85</td>
<td>42.88</td>
</tr>
<tr>
<td></td>
<td>Texture</td>
<td>Mixed</td>
<td>738.76</td>
<td>78.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>575.17</td>
<td>44.00</td>
</tr>
<tr>
<td>Unilateral</td>
<td>Color</td>
<td>Mixed</td>
<td>449.83</td>
<td>68.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>350.91</td>
<td>43.71</td>
</tr>
<tr>
<td></td>
<td>Shape</td>
<td>Mixed</td>
<td>473.55</td>
<td>74.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>366.49</td>
<td>42.88</td>
</tr>
<tr>
<td></td>
<td>Texture</td>
<td>Mixed</td>
<td>474.75</td>
<td>78.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>379.94</td>
<td>44.00</td>
</tr>
</tbody>
</table>

Analysis of variance found a significant main effect of response mode (F(1,54) = 37.83, MSE = 4155922.16, p < .001). Movement latency was significantly lower for the unilateral condition, where subjects had the possibility of initiating movement before determining the answer to the current trial. In addition to the main effect of response mode, movement latency was found to significantly vary according to context (F(1,54) = 42.02, MSE = 1237199.67, p < .001): subjects waited longer before initiating movement in the mixed block compared to the pure block. Figure 6-3 plots movement latency as a function of response mode and context.
Figure 6-3. Initial movement latency as a function of context and response mode. Error bars indicate 95% confidence intervals.

The fact that subjects initiated movement sooner (smaller latency) in the unilateral condition provides evidence that they were able to coordinate the cognitive object discrimination task and the motor act of reaching towards the waypoint. Latencies were higher in the mixed blocks of both the unilateral and bilateral conditions. It might be expected that in the unilateral condition, no effect of task context would be found since subjects could potentially begin movement as soon as the stimuli were displayed and could delay attending to the stimuli until after movement was initiated.

Such a strategy, while possible, is not necessarily optimal in the task environment since the arrival time of the mouse at the waypoint circle needs to be coordinated with the cognitive decision process. By initiating movement immediately in the mixed block, subjects risk arriving at the waypoint before a decision has been reached and consequently incurring a timeout error. The two possible means of avoiding this error are waiting longer to initiate movement in the mixed block, or alternatively initiating movement immediately but with a lower movement velocity. The fact that an effect of context was found in the unilateral condition suggests that at least on some portion of trials, subjects adopted the strategy of delaying their initial movement depending on the block context. However, even these delayed movements were initiated sooner than in the bilateral condition.
Figure 6-4. Empirical histogram of movement latencies for each response mode and block context. The bin size used for the histogram is 25 ms.

Figure 6-4 shows the empirical histograms of initial movement latency for each response mode and block context (mixed versus pure), computed using bins of 25 ms. It is immediately apparent from the figure that there is a large qualitative difference between the latency distributions for the bilateral and unilateral conditions. Movements were not only initiated much sooner in the unilateral condition than in the bilateral condition, but they were also executed with much less variance in their timing. The unusually sharp peak around 275 ms for the unilateral latency distributions suggests that on a significant number of trials, subjects began moving the mouse before the task stimuli were even displayed.

The way the experimental software was coded, the mouse cursor was hidden at the start of each trial. When the task stimuli were displayed, the mouse cursor was ‘warped’ to the center of the start circle, after which the cursor was revealed; however, there is no way to prevent subjects from having a nonzero hand velocity at this point in the trial. When the mouse cursor reached the edge of the start circle, movement onset was detected. The sharp peak in the unilateral latency distribution is likely due to the absence of variability due to the cognitive component of deciding how to answer the current trial, as well as the absence of variability due to motor planning and initiation.
6.4.6 Movements made in the wrong direction

In analyzing the data, it was found that there were instances on which subjects moved the mouse cursor in the opposite direction from the appropriate waypoint for a given trial. For example, if the correct waypoint on a trial was to the right of the start circle, and the minimum $x$ coordinate of the mouse cursor was to the left of the start circle, then the subject had moved the mouse to in the wrong direction. The occurrence of such wrong-way movements is analyzed in this section.

Overall, wrong-way movements occurred on 10.1% of trials. As might be expected, they were more frequent in the bilateral response mode, and were practically nonexistent in the unilateral response mode, as in this condition the initial movement was always in the same direction regardless of the correct response. Table 6-4 reports the proportion of such ‘wrong-way movements’ for each task condition.

**Table 6-4. Proportion of wrong-way movements by task condition.**

<table>
<thead>
<tr>
<th>Response mode</th>
<th>Task</th>
<th>Context</th>
<th>Mean</th>
<th>+/- 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilateral</td>
<td>Color</td>
<td>Mixed</td>
<td>0.19</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Pure</td>
<td></td>
<td>0.17</td>
<td>0.03</td>
</tr>
<tr>
<td>Shape</td>
<td>Mixed</td>
<td></td>
<td>0.20</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Pure</td>
<td></td>
<td>0.18</td>
<td>0.03</td>
</tr>
<tr>
<td>Texture</td>
<td>Mixed</td>
<td></td>
<td>0.24</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Pure</td>
<td></td>
<td>0.19</td>
<td>0.03</td>
</tr>
<tr>
<td>Unilateral</td>
<td>Color</td>
<td>Mixed</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Pure</td>
<td></td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>Shape</td>
<td>Mixed</td>
<td></td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Pure</td>
<td></td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>Texture</td>
<td>Mixed</td>
<td></td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Pure</td>
<td></td>
<td>0.01</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Analysis of variance on proportion of movements made in the wrong direction found that the effect of response mode was significant ($F(1,54) = 137.24$, $MSE = 2.90$, $p < .001$). Subjects initiated movement in the wrong direction more frequently in the bilateral condition. In addition, the main effects of context and task were also significant (for context, $F(1,54) = 6.61$, $MSE = 2.440E-02$, $p < .05$; for task, $F(2, 108) = 8.86$, $MSE = 1.074E-02$, $p < .001$). Wrong-way movements were more frequent in the mixed context, and were more frequent for the texture discrimination trials than for color or shape trials. Both of these effects are consistent with previous analyses demonstrating
higher incorrect response rates for these conditions. The interactions of context and task with response mode were also significant (both $p < .001$), which is to be expected given the virtual absence of wrong-way movements in the unilateral condition.

Subsequent analyses deal with properties of the actual movement from the start circle to the waypoint. As wrong-way movements would be expected to exhibit anomalous movement properties (in terms of the spatial path, duration, and velocity profile of the movement), trials exhibiting wrong-way movements are excluded from the analyses that follow. Thus, for the analyses that follow, three exclusion criteria were applied to the data. Trials with an incorrect response were excluded, as were trials where the subject missed the waypoint, and trials in which the subject moved in the opposite direction of the correct waypoint. Combined, these three criteria excluded 24.84% of trials.

6.4.7 Total trial time

Total trial time was defined as the time from when the task stimuli were displayed and the mouse cursor was unhidden to the moment that the mouse cursor entered one of the two response circles. The mean total trial time for each experimental condition is reported in Table 6-5.

**Table 6-5. Mean total trial time for each experimental condition.**

<table>
<thead>
<tr>
<th>Response mode</th>
<th>Task</th>
<th>Context</th>
<th>Mean</th>
<th>+/- 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilateral</td>
<td>Color</td>
<td>Mixed</td>
<td>1234.57</td>
<td>98.98</td>
</tr>
<tr>
<td></td>
<td>Pure</td>
<td></td>
<td>1124.07</td>
<td>79.13</td>
</tr>
<tr>
<td>Shape</td>
<td>Mixed</td>
<td></td>
<td>1285.51</td>
<td>124.09</td>
</tr>
<tr>
<td></td>
<td>Pure</td>
<td></td>
<td>1102.61</td>
<td>86.51</td>
</tr>
<tr>
<td>Texture</td>
<td>Mixed</td>
<td></td>
<td>1312.94</td>
<td>122.11</td>
</tr>
<tr>
<td></td>
<td>Pure</td>
<td></td>
<td>1096.86</td>
<td>87.88</td>
</tr>
<tr>
<td>Unilateral</td>
<td>Color</td>
<td>Mixed</td>
<td>1118.21</td>
<td>98.98</td>
</tr>
<tr>
<td></td>
<td>Pure</td>
<td></td>
<td>933.64</td>
<td>79.13</td>
</tr>
<tr>
<td>Shape</td>
<td>Mixed</td>
<td></td>
<td>1170.30</td>
<td>124.09</td>
</tr>
<tr>
<td></td>
<td>Pure</td>
<td></td>
<td>962.92</td>
<td>86.51</td>
</tr>
<tr>
<td>Texture</td>
<td>Mixed</td>
<td></td>
<td>1174.93</td>
<td>122.11</td>
</tr>
<tr>
<td></td>
<td>Pure</td>
<td></td>
<td>1011.95</td>
<td>87.88</td>
</tr>
</tbody>
</table>

Analysis of variance revealed significant main effects of response mode, task, and context. For response mode, trials were completed significantly faster in the unilateral
condition (F(1,54) = 4.55, MSE = 1436436.50, p < .05) compared to the bilateral condition. This effect shows that subjects were able to exploit the possibility of coordinating cognitive and motor processes in the unilateral response condition. For both response modes, trials in mixed blocks were completed more slowly than trials in pure blocks (F(1,54) = 39.44, MSE = 2643562.63, p < .001), demonstrating that the experimental manipulation of the block context significantly altered the difficulty of the task, both in terms of error rates but also trial completion time. The effect of trial task was also significant (F(2,108) = 8.13, MSE = 61406.88, p = .001). On average, color discrimination trials were completed the fastest (mean = 1102.6 ms), followed by shape (mean = 1130.33 ms) and texture trials (mean = 1149.17 ms). The difference between shape and texture trials was not significant. Thus, the difference in the difficulties of the three tasks were much smaller than the difficulty imposed by executing them in a mixed versus pure block context.

Learning over the course of the experiment was investigated by comparing mean trial completion time over groups of 20 trials. Figure 6-5 shows the speedup in trial completion time over the course of the experiment. Performance improved rapidly over the first 100 trials and then remained relatively stable over the remainder of the experiment. Since the order of mixed and pure blocks was counterbalanced across subjects, the presence of learning over the first portion of the experiment would not bias the results of the analyses.
A repeated-measures ANOVA on mean trial completion time using the successive groups of 20 trials as a factor revealed that the improvement in performance over the course of the experiment was significant (F(32,1760) = 12.44, MSE = 757105.60, p < .001).

6.4.8 Movement speed

After the stimuli were displayed and the subject initiated movement, the task required that subjects move the mouse from the start circle, through a waypoint, and into the appropriate response circle. Based on the analyses of movement latency, it was demonstrated that subjects adaptively altered the duration they waited before initiating motor movement from the start circle to the waypoint. It is possible that subjects altered not only the movement latency, but also the speed of the physical movement. Average movement speed was computed as the pathlength of the movement divided by the movement duration, and is reported in units of pixels per second. Movement speed is a more sensitive measure than movement duration, as it takes into account any trial-to-trial variation in the pathlength of the movement. By contrast, differences in movement duration could be due to differences in movement speed, or due to variation in movement pathlength (for example, adopting a more circuitous movement). Movement
pathlength is analyzed separately in the next section. The observed mean movement speed for each experimental condition is given in Table 6-6.

Table 6-6. Mean movement speed by experimental condition.

<table>
<thead>
<tr>
<th>Response mode</th>
<th>Task</th>
<th>Context</th>
<th>Speed (pixels/s)</th>
<th>+/- 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilateral</td>
<td>Color</td>
<td>Mixed</td>
<td>1081.67</td>
<td>99.71</td>
</tr>
<tr>
<td></td>
<td>Pure</td>
<td></td>
<td>1090.19</td>
<td>98.53</td>
</tr>
<tr>
<td></td>
<td>Shape</td>
<td>Mixed</td>
<td>1085.10</td>
<td>101.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>1129.09</td>
<td>104.54</td>
</tr>
<tr>
<td></td>
<td>Texture</td>
<td>Mixed</td>
<td>1081.73</td>
<td>98.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>1137.80</td>
<td>115.22</td>
</tr>
<tr>
<td>Unilateral</td>
<td>Color</td>
<td>Mixed</td>
<td>891.71</td>
<td>99.71</td>
</tr>
<tr>
<td></td>
<td>Pure</td>
<td></td>
<td>1006.20</td>
<td>98.53</td>
</tr>
<tr>
<td></td>
<td>Shape</td>
<td>Mixed</td>
<td>870.65</td>
<td>101.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>985.31</td>
<td>104.54</td>
</tr>
<tr>
<td></td>
<td>Texture</td>
<td>Mixed</td>
<td>850.35</td>
<td>98.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pure</td>
<td>1000.21</td>
<td>115.22</td>
</tr>
</tbody>
</table>

Analysis of variance revealed a significant interaction between context and response mode (F(1,54) = 5.70, MSE = 170646.40, p < .05). This interaction is plotted in Figure 6-6. In the bilateral condition, movements were executed with the same speed regardless of whether they occurred in a mixed block or pure block, while in the unilateral condition, movement speed depended on task context: the movement from the start circle to the waypoint was executed faster in the pure blocks compared to the mixed blocks. The significant interaction demonstrates that not only was movement initiation adaptively altered based on the task properties, but the speed of movement was also under adaptive control. The lack of an effect in the bilateral condition suggests that the difference in movement speed does not reflect mandatory interference due to the combination of cognitive and motor tasks, but rather an effect that only occurs when it is beneficial to task performance (in the unilateral condition), suggesting that it represents an adaptive strategy rather than interference.

The main effect of response mode was also significant (F(1,54) = 6.18, MSE = 2338642.64, p < .05): movements were executed faster in the bilateral condition than in the unilateral condition. This effect is consistent with a parallel execution strategy for the...
unilateral response mode, where the duration of movement must be adaptively coordinated with the temporal dynamics of the cognitive component of the task.

![Graph showing mean movement speed by context and response mode. Error bars indicate 95% confidence intervals.](image)

Figure 6-6. Mean movement speed by context and response mode. Error bars indicate 95% confidence intervals.

Figure 6-7 shows the empirical distributions of movement speed for each context and response mode. The histograms were computed using a bin size of 50 pixels/s. Compared to the distributions of movement latencies, the movement speed distributions are much more qualitatively similar. In the unilateral condition, the distribution is shifted to the left for mixed trials, indicating slower movement speed in this condition. This again is consistent with the hypothesis that subjects coordinated the execution of the motor movement with the cognitive task in the unilateral condition. For the bilateral condition, there is virtually no difference between the mixed and pure distributions, in terms of both the mean speed and the variance of the distributions, suggesting that subjects adopted a common serial processing strategy for all trials in this condition.
6.4.9 Movement pathlength

An ANOVA was conducted to determine whether the mouse pathlength significantly differed among any of the experimental conditions. Mouse pathlength was computed by summing the incremental Euclidean distance traveled by the mouse from the start circle to the waypoint based on the raw mouse data. Although the task was designed to impose constant movement difficulty across both response modes in terms of both distances and target sizes, it is possible that subjects could have deliberately adopted more circuitous movements; such a strategy would effectively slow down movements, and depending on the condition could be an adaptive strategy. For example, in the unilateral condition, subjects might adopt a longer movement by deviating upward or downward from the straight line path between the start and waypoint circles. However, based on the results of the ANOVA, there were no significant differences in mouse pathlength across any of the experimental conditions (all \( p > .05 \)).

6.4.10 Movement duration from the waypoint to response circle

After reaching the waypoint, the task required that subjects move the mouse into one of the two response circles within the 500 ms timeout period. The mean movement time from the waypoint to the response circle was 221.66 ms (SE = 4.38), well under the time
constraints imposed by the task. An analysis of variance found no significant differences in the duration of this movement as a function of response mode or task context (all \( p > 0.05 \)), suggesting that all subjects in all conditions had determined their response to the current trial by the time they reached the waypoint. However, an interaction between task and response mode was marginally significant (\( F(2,108) = 3.19, \text{MSE} = 653.671, p = 0.045 \)), as was the main effect of task type (\( F(2,108) = 3.08, \text{MSE} = 631.12, p = 0.05 \)). The interaction between task and response mode is reported in Table 6-7, and stems from a small but significant difference in the movement time for texture discrimination trials. The movement time for texture discrimination trials was slightly higher in the unilateral condition than in the bilateral condition, whereas the movement time did not differ for the color or shape trials. Since the magnitude of this difference was small (12.67 ms), it will not be considered further without replication of the results.

### Table 6-7. Interaction between condition and task for movement duration from the waypoint to the response circle.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Task</th>
<th>Duration (ms)</th>
<th>+/- 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilateral</td>
<td>Color</td>
<td>221.36</td>
<td>12.48</td>
</tr>
<tr>
<td></td>
<td>Shape</td>
<td>219.44</td>
<td>12.26</td>
</tr>
<tr>
<td></td>
<td>Texture</td>
<td>213.26</td>
<td>13.63</td>
</tr>
<tr>
<td>Unilateral</td>
<td>Color</td>
<td>227.15</td>
<td>12.48</td>
</tr>
<tr>
<td></td>
<td>Shape</td>
<td>222.80</td>
<td>12.26</td>
</tr>
<tr>
<td></td>
<td>Texture</td>
<td>225.94</td>
<td>13.63</td>
</tr>
</tbody>
</table>

#### 6.4.11 Velocity profiles for the initial movement

The smoothed mouse trajectories from the start circle to the waypoint were sampled at 5% quantiles of the movement duration, resulting in 21 measurements from each velocity profile (ranging from 0 to 100% in 5% increments). The previous analysis of movement speed demonstrated that average velocity significantly varied according to response mode and context. By examining the entire velocity profile it is possible to determine where and when in the course of the movement these effects were manifest. There are several possibilities; one is that all subjects, regardless of condition, initiated
movements at the same velocity\textsuperscript{19} and then sped up the movement once they had determined the answer to the current trial. This strategy would be reflected as a divergence in the velocity of each condition from a common starting velocity. By contrast, another possibility (and the hypothesis under investigation) is that subjects adaptively planned their motor execution based on an anticipation of the processing dynamics of the cognitive task. This strategy would be reflected in the velocity profiles that differ even at the onset of movement and persist throughout the entire trajectory. That is, subjects would initiate a movement with a velocity that depended on the expected cognitive processing dynamics for the current context.

Figure 6-8 shows the mean velocity profiles for each context and response mode. The top panel displays the velocity profiles as a function of the percent of the movement duration. As can be seen, there are differences in movement velocity even at the earliest measurements that depend on both the response mode and the block context. This is particularly noteworthy, as in the unilateral condition subjects initiated movement before they had determined the answer to the current trial. Thus, the difference in initial movement velocity depending on mixed versus pure blocks in the unilateral condition is indicative of subjects anticipating the temporal demands of the cognitive task. In the bottom panel, the data have been temporally aligned according to the mean movement latency and duration for each condition, such that the origin reflects the moment the task stimuli were displayed, and the start of each trajectory corresponds to the mean time of movement onset.

\textsuperscript{19} Movement onset was determined at the moment the mouse cursor left the start circle; thus the measured movements did not start at zero velocity.
Figure 6-8. Velocity profiles for the hand movement from the start circle to the waypoint, by experimental condition. (a) Velocity as a function of percentage of the total movement duration. (b) Temporally aligned mean velocity profiles.

In the unilateral condition, subjects initiated movement sooner, but with lower velocity, compared to the bilateral condition. In addition, movements were initiated slower, and later for the mixed trials compared to the pure trials. This difference in velocity between pure and mixed trials was primarily limited to the unilateral condition. By contrast, in the bilateral condition, subjects initiated movements later for mixed trials,
but did not vary (or only slightly varied) the movement velocity depending on context (mixed versus pure).

An additional distinctive feature of the velocity profiles is their sinusoidal shape. For all four conditions, velocity initially increased to a peak, decreased to a minimum, and began increasing again before the mouse cursor reached the waypoint. Peak velocity occurred at approximately 25% of total movement duration, while the minimum velocity occurred at about 85% of total movement duration. This pattern differs significantly from velocity profiles obtained in single point-to-point reaching movements, where the velocity profile is typically unimodal and bell-shaped (Harris & Wolpert, 1998; Todorov, 2004). This suggests that subjects did not approach the experiment as consisting of two sequential and independent motor tasks (moving from the start circle to the waypoint, then from the waypoint to the response circle), but rather planned and executed their movements by incorporating information regarding both the immediate and subsequent reaching targets.

6.4.12 Vertical deflections in movement trajectories

If subjects planned their trajectories to include not just the immediate waypoint, but the subsequent response, then it is reasonable to assume that the identity of that response (same versus different) should factor in the executed movement trajectory. Particularly relevant is the timing and extent to which subjects deflected their movements away from a perfectly horizontal line between the start circle and the waypoint. Given the time constraints imposed by the timeout period, it was hypothesized that for movements ultimately terminating on a target located in the upper half of the screen, subjects would deflect their initial movement towards the bottom of the screen. Doing so would enable subjects to execute a straight-line movement from the waypoint to the response circle, thus minimizing the time required to execute this movement.

To address this measure, vertical deflection was computed as the difference between the y-coordinate of the smoothed mouse trajectory and the horizontal line connecting the start and waypoint circles. The mean vertical deflection was then computed separately for each combination of response mode, context, and trial response (same vs. different). This resulted in 8 different conditions. Further, deflection was computed for each
quantile along the movement, from 0 to 100% in 5% increments. These data are shown in Figure 6-9. Dashed lines in the figure correspond to data from the bilateral condition, while solid lines are from the unilateral condition. Triangles correspond to trials where the subject responded ‘same’, and circles indicate responses of ‘different’. Filled markers are data from pure blocks, while unfilled markers are from mixed blocks.

![Figure 6-9](image)

**Figure 6-9.** Vertical deflection in mouse movements from the start circle to the waypoint. Triangles indicate trials where the response was 'same', circles correspond to responses of 'different'. Filled markers indicate pure blocks, while open markers are data from mixed blocks. Solid lines are from the unilateral condition, dashed lines are from the bilateral condition.

The most obvious feature of the data in Figure 6-9 is the difference in deflection between the unilateral and bilateral conditions. In the bilateral condition, subjects consistently deflect their mouse trajectories downward\(^{20}\). This feature is to be expected,\(^{20}\)

\(^{20}\) Downward is defined in screen coordinates as deflection towards the bottom of the display. In the physical movement space in the horizontal plane, downward deflections correspond to moving the hand closer to the body.
given that the final target in the bilateral condition was always in the upper half of the display. By deflecting the initial movement downward, subjects could execute a straight line movement from the waypoint to the response circle, and thus minimize the transit time. The negative deflection is apparent in the bilateral condition even at the detection of movement onset.

By contrast, in the unilateral condition, movements were initiated horizontally, with little or no deflection present. The direction of eventual deflection depended on the response that was to be made on that trial: for ‘same’ trials, subjects deflected their movement downward, while they deflected their movement upwards for ‘different’ responses. Again, this pattern is consistent with an anticipatory strategy to facilitate the execution of the second half of the motor task. In the bilateral condition, there is a small difference (~2 pixels) between same and different responses. Responses of ‘same’ exhibited slightly less negative deflection than different responses. A possible reason for this is that in the bilateral condition, the ‘same’ response circle was located in the left half of the display. The decreased magnitude of deflection may be due to the fact that this condition required subjects to move the mouse from the lateral side of the workspace towards the sagittal plane, rather than vice versa. Small differences in the curvature of reaching movements have previously been reported as a function of the location and direction of movements in the physical workspace. In particular, the hand has been shown to have varying inertia depending on the direction of movement (Gordon, Ghilardi, Cooper, & Ghez, 1994).

The emergence of response-contingent deflection in the unilateral condition appears early in the movement, at approximately 25% of the total movement duration. This suggests that subjects in the unilateral condition determined their eventual response relatively early in the movement (well before they reached the waypoint). This, combined with the fact that movements were initiated both sooner and slower in the unilateral condition compared to the bilateral condition, suggests that subjects adapted all aspects of the motor plan based on the ongoing properties of the task, rather than merely delaying the execution of stereotyped, “pre-canned” movements.
6.4.13 Summary and discussion of results

This experiment was designed to test the hypothesis that subjects would learn to intricately coordinate properties of their motor systems depending on the dynamics of ongoing cognitive processing. This was examined by implementing a paradigm (the ETA paradigm) where the difficulty of simple geometric object discrimination varied as a function of pure versus mixed blocks of tasks. Previous research has shown that frequent task switches impair cognitive performance in these sorts of simple decision tasks, resulting in increased response time as well as increased occurrence of errors. In addition to the cognitive decision component, the ETA paradigm included a motor task of moving from a start circle, through a waypoint and into an appropriate response circle. Although subjects were free to adopt any processing strategy to complete each trial, the evidence suggests that in the unilateral response mode, subjects adaptively controlled numerous low-level properties of their motor systems to maximize performance in the task depending on the cognitive dynamics of the object discrimination task.

In particular, in the unilateral condition subjects exhibited a lower response latency compared to subjects in the bilateral response mode. This effect is particularly important, as it suggests that subjects initiated movement before they had determined the appropriate response to the current trial. Concomitant with their lower latency, subjects in the unilateral condition exhibited slightly higher error rates (the overall error rate was 2.67% higher).

The lower latency in the unilateral condition extended to a faster overall trial completion time for this condition. Note that the two response modes were of equivalent difficulty; both placed the same cognitive and motor demands on the subjects, and subjects were free to adopt a serial or sequential processing strategy in both conditions. The fact that subjects in the unilateral condition completed each trial significantly faster demonstrates that they were able to coordinate the cognitive and motor tasks.

Detailed analysis of the motor trajectories revealed that not only did subjects initiate movement more quickly in the unilateral condition, they also adaptively tuned the properties of their motor trajectories based on the ongoing task context (defined by mixed versus pure blocks of trials). In the mixed block, subjects executed movements
more slowly compared to the pure block. This effect disappeared in the bilateral condition. Importantly, differences in movement speed appeared even at the detected onset of movement in the unilateral condition. This suggests that the motor program was selected or tuned based on the expected time required to complete the cognitive decision task, rather than being merely due to subjects speeding up or slowing down once they had determined the answer to the current trial. In addition, the speed of the movement was significantly slower overall for the unilateral condition than for the bilateral condition.

In sum, the evidence provided by this experiment strongly argues in favor of the main hypothesis of this thesis. When given the opportunity, humans will acquire and exploit predictive models of their ongoing performance capacities and dynamics. As was previously demonstrated with the Bayes’ Ball task, these internal models predict not just properties of the perceptual and motor systems, but also detailed properties of ongoing cognitive processing—be it the uncertainty in temporal estimation, or the time required to discriminate between objects in a task switching paradigm. While the Bayes’ Ball paradigm heavily constrained the space of possible strategies available to the subject, the ETA paradigm deliberately allowed for a wide range of possible strategies. Given this freedom, the evidence suggests that subjects were able to adaptively exploit it to their performance benefit.

### 6.5 Minimum jerk model of the via-point reaching task

The minimum jerk model (Flash & Hogan, 1985; Hogan, 1984) is a widely influential model of motor planning in simple reaching tasks. According to the model, complete motor trajectories are planned offline (before execution), based on the optimization of a task-invariant cost criterion. Once the motor plan has been specified, it is up to the motor system to execute the pre-specified plan as accurately as possible. The popularity of the minimum jerk model stems in part from its ability to explain the relatively straight end-effector (hand) paths and bell-shaped velocity profiles commonly

---

21 As of 11/17/08, (Hogan, 1984) had garnered 403 citations, and (Flash & Hogan, 1985) had received 862 citations on Thomson ISI Web of Science.
observed in point-to-point reaching movements. Another reason for its popularity however, is its ability to account for human performance with a minimal number of parameters and using a task-independent organizational principle: namely, maximizing the smoothness of the movement.

Flash and Hogan (1985) derived a version of the minimum jerk model for the special case of moving the hand from a starting position, through a via-point, to a final target. However, their model assumed that the hand was required to stop at the final target (zero velocity and acceleration). Although subjects in the ETA experiment were not given any constraints regarding their final velocity or acceleration, all subjects adopted a strategy of moving through the response circle. Stopping accurately at a target is a more time-consuming motor task than hitting or moving through it (Liu & Todorov, 2007), and given the 500 ms timeout period imposed on subjects, the observed strategy of moving through the response circles is an adaptive response.

In order to test the predictions of the minimum jerk model for the ETA experiment, a modified version of the minimum-jerk model was derived for the special case where the final hand velocity and acceleration are unconstrained. The derivation for this model is provided in Appendix C.

The modified minimum jerk model was used to generate predicted velocity profiles for the complete motor movement (from the start circle to the response circle). The exact distances and positions from the human experiment were used to define the location of the waypoint and targets. Since the minimum jerk model requires that the total movement duration be specified as input, the observed mean total movement durations (movement from the start circle to the response circle) from each response mode and context in the experiment were used. One complication is that the time of physical movement onset for the human subjects was not directly recorded, but only the time at which the mouse cursor left the start circle (which had a radius of 15 pixels). Therefore, the total movement duration specified as input to the minimum jerk model was estimated as the empirically observed movement duration, plus an estimate of the time required to move the mouse 15 pixels. This was estimated to be 100 ms, based on a previous estimate by (Liu & Todorov, 2007) for the time required to move the hand 1 cm from a
complete stop. Subsequent investigation showed that the qualitative properties of the resulting trajectories did not depend on the exact value used.

The purpose of applying the modified minimum jerk model to the human data was twofold. First, a main hypothesis of the ETA experiment is that properties of motor planning and execution are not separate processes from cognition or perception, but rather are coordinated with the embodied dynamics of a task, including in this case the cognitive dynamics of discriminating between objects. By contrast, the minimum jerk model assumes that all motor movements are planned independently of cognitive and perceptual processes, by optimizing a fundamentally task-independent criterion (maximizing the smoothness of the movement). If the human reaching trajectories qualitatively differ from the predictions of the model, then it can be safely concluded that humans were not, as their primary objective, maximizing movement smoothness in the experiment. The second objective of comparing the minimum jerk model to the empirical data is to determine whether the differences in the motor trajectories in the different experimental conditions can be attributed purely to differences in total movement duration. That is, do the differences in velocity profiles between conditions automatically fall out as a consequence of first deciding on a total movement duration? This question is relevant as it provides insight into the number of control variables that humans exploited in the experiment in order to maximize their performance.

The predicted trajectory generated by the minimum jerk model is shown by the solid curve in Figure 6-10a. The waypoint passage time $t_1$ was optimized to maximize the smoothness of the overall movement. The completion time $t_f$ for the movement was specified based on the empirical movement time for the bilateral mixed condition. The behavior of the minimum jerk model is invariant to rotations or translations of the targets (Flash & Hogan, 1985), consequently the same trajectory (appropriately mirrored) would result for movements to the left or right, or to an upper or lower response circle. It should be noted however, that human reaching movements are not completely invariant to these same transformations (Engelbrecht, 2001; Gordon et al., 1994).
Figure 6-10. Comparison of minimum jerk trajectories (solid lines) and empirical trajectories (dashed lines). (a) Minimum jerk trajectory computed with waypoint passage time optimized to maximize movement smoothness. (b) Minimum jerk trajectory with waypoint passage time set to the empirically observed value. Empirical data is based on the bilateral mixed condition.

The dashed curve in Figure 6-10a shows the empirically observed trajectory for ‘different’ responses in the bilateral mixed condition. This condition was chosen for comparison as it is assumed that human subjects had already determined their final response for the current trial before initiating movement, and because it was the group exhibiting the largest vertical deflection in the human data (reaching a maximum vertical deflection of approximately 15 pixels). It is immediately apparent from the figure that the magnitude of deflection for the minimum jerk model is much greater than that observed in the human data. For the minimum jerk model, maximum deflection reached 50 pixels, a magnitude over $3 \times$ greater. It is clear from the large difference between the predicted and observed trajectories that humans were not using movement smoothness as the planning criteria for their movements in the task.

In Figure 6-10b, the movement time between waypoint passage and the response circles was set to the empirically observed mean value, $t_f - t_i = 221.66$ ms. As reported previously, this value did not significantly differ between response modes or block contexts. It can be seen in Figure 6-10b that the movement trajectory is much more accurately captured when this additional constraint is placed on the minimum jerk model. It should be noted however, that the model’s predictions no longer represent the minimum jerk trajectory. Instead, the model is adopting the human strategy, which presumably reflects the desire to reach the final target before the timeout period, while
maintaining an acceptable amount of accuracy. How the values $t_i$ and $t_f$ are selected by the motor planning system is outside of the theory.

In addition to predicting the position of the hand over time, the minimum jerk model also provides predictions regarding the velocity and acceleration of the hand over the course of the movement. The equations for velocity and acceleration are provided in Appendix C. Based on the $x$ and $y$ velocities, tangential velocity of the hand is computed as

$$v(t) = \sqrt{x'(t)^2 + y'(t)^2}.$$

Figure 6-11 compares the empirically observed velocity profiles to those predicted by the modified minimum jerk model, where $t_i$ and $t_f$ were constrained to the empirically observed values. Movement onset for the minimum jerk model was defined as the time at which the predicted hand position had moved 15 pixels from the starting location (corresponding to the exit-time from the start circle in the human experiment). The large discrepancy between the observed and predicted velocity profiles is immediately apparent. Compared to the human data, the model initiates movements with a much lower velocity. Note that the minimum jerk model would be capable of exhibiting the same initial velocity as human subjects, as there is no constraint placed on the magnitude of the control signal. However, such a velocity profile would not result in a maximally smooth movement. The predicted velocity is bell-shaped, reaching a maximum at around 75% of the movement duration. By contrast, human subjects initiated movement with a much higher velocity, and the profiles exhibit a distinct sinusoidal shape. This occurred even in the bilateral response mode, where there is strong evidence that subjects had determined the motor target before initiating movement.
Figure 6-11. Comparison of empirical and predicted velocity profiles. (a) Human data, re-plotted from Figure 6-8. (b) Velocity profiles predicted by the minimum jerk model for each experimental condition, with waypoint passage time and total movement time constrained to the empirically observed values.

Since the performance objective of the model is explicitly specified, any deviations between human performance and model predictions can be definitively characterized as decreasing the smoothness of the movement. The large deviations between the model and empirical data suggest that maximizing the smoothness of movement to the exclusion of all else was not a likely performance objective adopted by the human
subjects in this experiment. The ETA paradigm imposed penalties associated with missing the via-point or target that are not captured by the task-independent minimum jerk model. More importantly, the large differences in initial movement velocity for the pure versus mixed blocks are absent for the predicted velocity profiles, even when the parameters $t_1$ and $t_f$ were chosen to fit the human data. For the human data, the difference in velocity between pure and mixed blocks is large at the earliest measurements of velocity and persists throughout the trajectory. For the minimum jerk model a different pattern holds: differences in movement velocity between pure and mixed blocks are smallest at movement onset, and increase slowly over the course of the movement. This result demonstrates that the empirically observed large difference in initial movement velocity cannot be attributed to properties of the motor trajectory or its timing in isolation. Rather, the failure of the minimum jerk model to account for the properties of the human data lends support in favor of the argument that the empirically observed trajectories reflect the combined influence of the cognitive and motor tasks, rather than representing a solution to the pure motor task on each trial.

Finally, the qualitative differences between the unilateral and bilateral velocity profiles cannot be accounted for by the minimum jerk model. This demonstrates that the differences in velocity profiles in the different conditions of the human data cannot be simply attributed to differences in the total movement duration. That is, it is not the case that human subjects selected a total movement duration for each condition, and the observed velocity profile differences were automatic consequences of this initial decision. Instead, it appears that humans controlled not only the total movement duration, but also properties of the movement at the beginning, and throughout the trajectory. How these properties of the movement are selected, as well as the differences in latency before movement onset, are fundamentally outside the scope of the minimum jerk model, which assumes that hand trajectories are planned and executed independently of all ongoing cognitive or perceptual processing.

6.6 Elements of an ideal performer analysis

Given the failure of the minimum jerk model to capture qualitative properties of the observed movement and velocity profiles, it is worth considering what aspects of the
model led to its failure, and how an ideal performer model might be constructed for the ETA paradigm. By definition, the minimum jerk model assumes that humans will plan motor trajectories that maximize the smoothness of the resulting movement, at the exclusion of all other factors. But when humans are given a specific task to accomplish, the merit of maximizing movement smoothness above all else becomes somewhat questionable. Indeed, it is questionable why movement smoothness is a biologically relevant organizing principle for movements in any task (Engelbrecht, 2001). When viewed in this light, theories such as the minimum jerk model are simply concise characterizations of empirical findings in particular tasks. When the model captures the properties of movements, it has succinctly described them (but has not ‘explained’ them), and when it fails to reflect human performance, the model offers no insight as to why or how.

More recent theories of motor control have attempted to circumvent this issue by adopting the task to be performed as the explicit optimization principle. If the task is to move the hand as quickly as possible from one point to another with a specified level of accuracy, then the optimal control policy is the one that accomplishes exactly this. Constraints on the motor control system are specified in terms of the need to execute movements with a stochastic motor system, and to correct for errors in execution online using uncertain sensory information.

An early model adopting this approach is the iterative stochastic optimized submovement model (Meyer, Kornblum, Abrams, Wright, & Smith, 1988). According to this model, rapid spatially-constrained movements are executed with a series of successive ballistic submovements. This model is able to account for the speed-accuracy tradeoff in movements specified by Fitts’ Law. However, some assumptions of the model are overly simplistic with regard to more complex movement tasks, in particular its restriction to one-dimensional motion and intermittent control of ballistic submovements.

A more recent model known as the minimum variance model (Harris & Wolpert, 1998) is able to account for the same speed-accuracy tradeoff, using continuous rather than intermittent control. In this model, Fitts Law as well as the empirical emergence of smooth, bell-shaped movements can be predicted from the signal-dependent noise
characteristics of the motor system. Applying large forces has the consequence of decreasing the accuracy of the movement. By exerting the least control possible to accomplish a task in a given amount of time, this model has as a side effect producing the smooth velocity profiles that are typically observed. If accuracy or speed demands are altered, the model’s predictions reflect the revised task given to the motor system, rather than the emergent epiphenomenon of maximizing smoothness.

Finally, Todorov (1998, 2004, 2007; Todorov & Jordan, 2002) has developed a theory of feedback-based, continuous control based on results from stochastic optimal control theory. In addition to its emphasis on uncertain feedback and online control, the optimal feedback control framework applied to the motor system emphasizes that “cognitive expectations, goals, etc. can have a significant influence on all aspects of the movement, making a strict hierarchical separation very unlikely” (Todorov, 1998; p. 7).

This progression of theories in motor control can be seen as a general weakening of arbitrary and task-independent constraints (such as movement smoothness) in favor of a motor planning and execution system that is optimized for particular goals to be achieved in particular task settings. The definition of motor optimality depends on properties of the visual system as well as the task that must be performed. The philosophy of this approach is entirely consistent with the main focus of this dissertation; namely, the components of interactive behavior cannot profitably be understood in isolation from the rest of the cognitive system, as well as the task that is to be performed.

Unfortunately, with the increased generality of the optimal feedback control theory, the mathematical tractability of its predictions quickly becomes an issue. Analytic solutions are generally only obtainable for simple deterministic systems, or for stochastic systems characterized by linear system dynamics, quadratic cost functions, and additive Gaussian noise (this combination of properties is known as the LQG control framework). Problems that cannot be cast in the LQG framework can still be solved using a discretized state space via dynamic programming (Bertsekas, 2000), but this solution quickly runs into the curse of dimensionality.

At present, it is not known whether there is a simple analytic form to describe optimal feedback control in the ETA paradigm. Instead, a rather simplified version of the task will be formulated and solved in the framework of dynamic programming. Dynamic
programming necessitates the use of a discretized state space, as well as a simplified task to maintain computational tractability. The results of the model will not be used to quantitatively compare human performance in the task to optimal predictions, but rather to demonstrate that human performance exhibits the same qualitative features as those derived based on this analysis.

6.6.1 Optimal control problem formulation

Consider a one-dimensional reaching task, where the agent must move its hand from a start position to a terminal position as quickly as possible. This is intended to correspond to the initial hand movement from the start circle to the waypoint in the ETA task. The hand is modeled as a fully observable second-order linear system, where the position and velocity are observed and noise-free and the control signal corresponds to acceleration of the hand. Concurrent with the movement task, the agent is performing an abstracted cognitive decision making task. At some time $t_d$, the decision is made and available to influence the agent’s motor system. If the hand arrives at the target before the decision is reached, a penalty is incurred. This penalty loosely corresponds to the timeout error that could occur when human subjects reached the waypoint before determining the answer to the current trial. The agent is free to wait for the decision before initiating movement (as were human subjects), but this strategy is presumably not optimal, as it excludes the possibility of cognitive-motor coordination. Further, it is assumed that time until the decision is reached is a random variable, characterized by a minimum time $t_{\text{min}}$ and variability drawn from a geometric distribution$^{22}$ with success probability $p_{\text{success}} = 0.2$ on each time step. Likewise, the agent does not know exactly when the decision will be available, but rather only has a probabilistic internal model capturing the statistical properties of the decision-making process. If the agent has access to this internal predictive model, and knowledge of the dynamics of its own motor

$^{22}$ A geometric distribution is the discrete-time equivalent of the exponential distribution. The discrete-time version is appropriate since the analysis is conducted using a discretized state space.
system, what is the optimal control policy for the agent in this abstract decision making environment?

This problem can be formulated as a Markov decision process as follows. The internal state of the agent corresponds to its hand position, velocity, elapsed time since the trial started, and a binary variable indicating whether the cognitive decision is available. Actions correspond to hand accelerations. In addition, the agent has access to a probabilistic transition model $P_{s \rightarrow s'}^a$, which specifies the probability of observing state $s'$ by starting from state $s$ and executing action $a$. The only probabilistic component of this transition model is the binary cognitive decision variable: on each time step after the minimum decision time, there is a fixed probability, $p_{\text{success}} = 0.2$, that the decision will become available. The position and velocity transitions are deterministic, and are governed by the second-order system dynamics:

$$\begin{align*}
\dot{x}(t) &= v(t) \\
\dot{v}(t) &= u(t).
\end{align*}$$

Each trial is terminated when the hand reaches the target, or when a maximum time has elapsed. The cost function to be minimized is specified as

$$\text{cost} = t_f + \sum_{t=0}^{t_f} \left[ w_{\text{energy}} \cdot u(t)^2 \right] + \begin{cases}
5000 & \text{if decision} = 0, \\
0 & \text{otherwise}
\end{cases}$$

where $t_f$ is the time the hand reaches the target, $u(t)$ is the hand acceleration at time step $t$, and $w_{\text{energy}}$ is a weight parameter that favors minimal control strategies. The value of $w_{\text{energy}}$ was set to a small value, $1.0 \times 10^{-6}$, to emphasize task completion over smoothness of movement. The state space was discretized using

- $\text{time} = \{0, 1000\}$, step size = 50 ms;
- $\text{position} = \{0, 300\}$, step size = 1 pixel;
- $\text{velocity} = \{0, 1600\}$, step size = 8 pixels/s;
- $\text{acceleration} = \{-5600, 5600\}$, step size = 160 pixels/s$^2$.

Dynamic programming (Bertsekas, 2000) was used to derive the optimal control law for this task environment. The solution to this problem consists of the optimal cost-to-go for each of 2.4 million states. The optimal control policy is the action (hand acceleration)
that minimizes cost-to-go for every state. By including time as a state variable, the state transition matrix is ensured to be acyclic, and the problem was solved using a single sweep of the value iteration algorithm, proceeding backward in time from the maximum time of 1000 ms.

6.6.2 Optimal control model results

The ideal performer analysis was run twice, using two values of the minimum decision time parameter, \( t_{\text{min}} = \{300, 500\} \) ms. Recall that the decision time was modeled as a random variable drawn from a geometric distribution added to the minimum decision time. For the distribution parameter \( p_{\text{success}} = 0.2 \), this resulted in mean decision times of 550 ms and 750 ms. For each solution, 500 trials were simulated by starting at the initial position \( (x_0, v_0) = (0, 0) \) and following the optimal control policy to obtain velocity profiles of the hand. The velocity profiles were fit using the same cubic spline interpolation method as applied to the human mouse trajectories from the ETA experiment, and re-sampled at 5% quantiles of the total movement duration.

Figure 6-12. Optimal velocity profiles as a function of (a) percent of total movement duration, and (b) trial time, computed for two values of the minimum decision time parameter.

Figure 6-12 illustrates the optimal velocity profiles for this decision environment, for each value of \( t_{\text{min}} \). Although the task is somewhat simplified compared to the task performed by humans in the ETA experiment, the same pattern of qualitative features is
present in the data. Namely, the velocity profiles exhibit a distinct sinusoidal shape, with peak velocity occurring at around 25\% of total movement duration, and a local minimum in velocity occurring around 85\% of movement duration. The profiles differ for the two values of $t_{\text{min}}$, which approximately correspond to the differences in decision time between the pure and mixed conditions in the human experiment. Large differences are present in the velocity profiles throughout the movement, apart from an initial period of rapid acceleration. Figure 6-12b illustrates that these differences are present well before the minimum decision times. Note that these features are all present despite the simplified one-dimensional version of the reaching task, and the absence of motor noise or accuracy constraints on performance. Instead, the non-bell shaped velocity profiles and differences between conditions are emergent properties of the optimal control solution for the combined cognitive-motor task.

Like human subjects, the optimal control model was free to adopt a serial strategy; wait until the cognitive decision is available, then initiate the motor task. In the absence of a predictive model of the temporal dynamics of the cognitive process, this strategy may well be optimal. However, when given the opportunity to acquire or exploit an internal model of its own cognitive dynamics, the optimal solution to the task is to incorporate this knowledge in the planning and online execution of motor movements.
7. Summary and conclusions

7.1 Introduction

How are embodied cognitive, perceptual, and motor processes coordinated in a manner that is both efficient and relevant to the immediate tasks and goals facing the individual? The purpose of this dissertation has been to shed further light on the answer to this question. In particular, the dissertation raised and supported two principal hypotheses:

1) The coordination of interactive behavior is sensitive to the low-level dynamics and variability intrinsic to human cognitive, perceptual, and motor processes, and to the external task environment.

2) The cognitive system achieves this coordination in routine behavior by learning and exploiting internal models of these embodied task dynamics.

The approach used to investigate these hypotheses was conducted on the embodiment level of analysis (Ballard et al., 1997)—the level of analysis at which the constraints of the component processes of interactive behavior first have implications for the ability to achieve goals in the world. The embodiment level of analysis emphasizes that the low-level properties of these components can have non-trivial implications for the effectiveness of alternative strategies for achieving goals, and can offer insights into the mechanisms responsible for their coordination.

In addition to emphasizing the primacy of the embodiment level, the work reported in this dissertation has adopted the rational analysis framework (J. R. Anderson, 1990) as a tool for uncovering the mechanisms underlying skilled interactive behavior. Rational analysis is an approach to understanding human cognition that emphasizes the need to study the cognitive system and its external environment in tandem. By assuming that the cognitive system is highly adapted to perform well in its environment, researchers can develop theories by positing how the cognitive system should perform, rather than grasping at unprincipled accounts for how it does perform.

While rational analysis has traditionally focused on adaptation to broad statistical properties of the external world, this research has applied the same methodology to the internal components of interactive behavior. While tasks and goals may change rapidly, the elementary components of behavior (specific perceptual, motor, and cognitive
processes) are relatively invariant, though not constant, across tasks. This fact argues for the rationality and feasibility of acquiring internal models of these elementary processing components. By predicting the dynamics of ongoing performance, the cognitive system can more efficiently coordinate interactive behavior in routine tasks. In each of the experiments conducted in this dissertation, quantitative predictions derived based on the rational analysis approach were found to be in close agreement with the empirical findings of human performance.

The focus of this dissertation has been on the intelligence of routine interactive behavior, rather than higher-level capacities that might be considered more representative of human intelligence, such as language or complex reasoning. However, it is clear that even on tasks as basic as reaching one’s hand from one location to another, human performance surpasses the capabilities of all currently known algorithms. The intricate complexity of routine interactive behavior is hidden by its seemingly effortless execution, much as the complexity of the human visual system is obscured by the effortless subjective experience of our ability to ‘see’.

7.2 Summary

Chapter 2 of this dissertation provided a review of recent literature that has investigated the coordination of cognitive, perceptual, and motor processes in routine interactive tasks. The empirical findings reviewed in Chapter 2 represent the sort of phenomena that need to be accounted for by a theory of interactive behavior. These findings have previously been accounted for by a wide range of vague mechanisms, but lacked a unifying explanation, and more importantly lacked a formal theory detailed enough to be implemented as a running computational model that performs the relevant task. A contribution of this dissertation was to provide a unifying explanation in terms of adaptation to and exploitation of internal models of embodied dynamics. While details of this theory may eventually be found lacking or inaccurate, it has the virtue of being implemented, and consequently falsified.

Chapter 3 provided a formal definition for the construct of an internal model, and offered a computational exploration of its properties. Several algorithms that have previously been proposed in the literature on motor learning were implemented and
evaluated on a simple test problem. Each of these exhibited significant limitations on the simple test, demonstrating the limits of currently known learning algorithms for acquiring internal models based on self-guided exploration in a task. The adaptive internal model framework was also developed in this chapter, designed to illustrate how an internal model can not only be acquired, but also profitably used by the cognitive system to control and improve performance in routine interactive tasks. The structure and organization of this modeling framework was justified both by its functionality in biologically relevant tasks, as well as its general agreement with the known biology and computational function of the human cortex, cerebellum, and basal ganglia. While it is unlikely at this time that neuroscientific evidence can be used to definitively demonstrate or rule out a particular mechanism, incorporating this research provides a powerful additional constraint on the space of theoretical cognitive mechanisms.

The adaptive internal model framework was used to develop a cognitive model of a simple reaching task that has previously been characterized as evidence for an abstract metacognitive control system (Augustyn & Rosenbaum, 2005). This application not only demonstrated the computational sufficiency of the hypotheses of this dissertation, but also provided a quantitative, and potentially falsifiable account for performance in a task where only a vague explanation had previously been given.

Chapter 4 introduced Bayes’ Ball, a novel paradigm developed as a specific test of the hypotheses of this dissertation. In particular, this paradigm demonstrated that humans are able to acquire and exploit an internal predictive model of a rather subtle property of the ability to estimate time intervals. This property, known as the Weber law of interval timing (Gibbon, 1977; Staddon & Higa, 1999), is manifest in the linear increase in the standard deviation of time estimates as a function of the magnitude of the interval that is to be estimated. While this property has long been the study of psychologists, the approach of this dissertation was to examine the extent to which naïve subjects are implicitly sensitive to the implications of the Weber law in their behavior. The Bayes’ Ball paradigm was inspired by recent experiments (Maloney et al., 2007; Trommershäuser et al., 2006; Trommershäuser et al., 2003) designed to assess implicit knowledge of the uncertainty inherent in reaching movements. In the research reported
in this dissertation, the theoretical focus was not on internal models of motor dynamics, but rather the cognitive dynamics of temporal interval estimation.

An ideal performer analysis was conducted to determine whether human performance matched an optimal decision strategy, given the performance constraints imposed by the Weber law and features of the task environment. It was found that human subjects closely approximated optimal performance in the Bayes’ Ball task. Cognitive models based on the adaptive internal model framework were developed as an explicit test of the explanatory sufficiency of the main hypotheses of this thesis. It was shown that, with no free parameters, a cognitive model that acquires and exploits an internal model of its embodied dynamics was able to closely match human learning and asymptotic performance. When the model was extended with two parameters, it was also able to offer a plausible explanation for the nature of the small empirical deviations from optimal performance. An alternative model, one that learned based on acquiring a mapping from actions directly to their expected utility (using the computational approach of reinforcement learning), was unable to account for either the time course or asymptotic performance of human subjects in the experiment. Combined, the results of the experiment, ideal performer analysis, and cognitive models offered strong and converging evidence in favor of the human capacity to acquire and exploit internal models of embodied dynamics.

In chapter 5, a second experiment using the Bayes’ Ball paradigm was reported. A primary motivation for the experiment was to test a prediction made by the cognitive model developed for the first experiment. The cognitive model made the prediction that by altering the paradigm to use a blocked rather than randomized trial design, one of the observed biases in human performance would disappear. This prediction was tested in the second experiment, and supported by the empirical results. In addition, a between-subject manipulation tested whether subjects would adapt to an external source of variability, or whether the ability to acquire internal models is restricted to internal dynamics. This question is of some interest, as the philosophical approach of active externalism (Clark, 2003) claims that to the adaptive system, internal processes are no different than external processes. Human subjects adapted to the external variability as readily as their own internal variability, in keeping with the active externalism
hypothesis. The noise manipulation also illustrated the need to account for human learning, as it ruled out the possibility that human subjects came in to the experiment with a fixed decision strategy based on prior experience.

The design of the second experiment also enabled a strong test of whether subjects were acquiring internal models of their variability, or instead learning a mapping from time estimates to expected reward. For all subjects, trial 226 was the first trial on which asymmetrical penalties required shifting away from the center of the reward region. It was observed that on this trial, subjects shifted their mean aiming position in a manner adaptive to the penalty configuration, contra to the predictions of any theory based around maximizing scalar reinforcement. A post-experiment questionnaire demonstrated that subjects’ implicit performance in the task was more optimal than their explicit judgments of the optimal task completion strategy. When explicitly asked for the optimal strategy, subjects showed no awareness of the Weber law of interval timing, or of the artificial noise manipulation.

In keeping with the approach adopted for the first experiment, a combination of ideal performer analysis and cognitive modeling were used to determine the optimality of human performance, as well as a plausible means by which this performance could have been learned. In the second experiment, subjects’ task scores were equivalent to those based on the predictions of the ideal performer analysis. The cognitive model was able to predict this performance with no parameters adjusted to fit the human data. This provided an additional test of the computational sufficiency of the hypotheses of this dissertation.

Chapter 6 reported the results of a third experiment, using a paradigm referred to as ETA (short for estimated time of arrival). Compared to the first two experiments, the ETA paradigm provided relatively few constraints on the space of possible strategies that human subjects could adopt to complete the task. This was deliberate, as it was hypothesized that the acquisition of internal models of embodied dynamics reflects a general feature of routine interactive behavior, rather than a peculiarity that only emerges under highly restrictive task settings or with the deliberate intention to do so. The ETA paradigm combined a simple cognitive discrimination task with an equally simple motor reaching task. The difficulty of the cognitive decision task was
manipulated by the use of a task switching paradigm, and the key theoretical question was whether the integrated cognitive-motor system would exhibit behavior reflecting anticipation of and adaptation to the varying difficulty of the cognitive task.

The empirical results from the ETA experiment indeed supported this hypothesis. Although subjects were free to adopt any processing strategy to complete each trial, the evidence suggests that subjects adaptively controlled low-level properties of their motor planning and trajectories to maximize performance, depending on the cognitive dynamics of the object discrimination task.

Human subject performance in the ETA experiment was further analyzed by computing the predictions of a widely popular model of motor trajectory planning known as the minimum jerk model (Flash & Hogan, 1985; Hogan, 1984). An extension of the minimum jerk model was derived for via-point tasks where terminal velocity and acceleration of the hand are unspecified. The minimum jerk model predicts that motor reaching trajectories are planned and subsequently executed by maximizing the smoothness of the overall movement, independent of the actual goal that is to be achieved through the motor movement. By contrast, this dissertation hypothesized that motor control in routine interactive tasks would not occur independently of ongoing cognitive or perceptual processes, but would instead be intricately coordinated in order to maximize task-specific performance. In support of the hypotheses of this dissertation, the minimum jerk model failed to capture even qualitative properties of the empirically observed trajectories and velocity profiles.

As an alternative explanation, an optimal control model was developed that explicitly optimized task-specific performance by incorporating an internal model of the dynamics of the cognitive decision process into motor planning and online execution. This optimal control model was cast in the framework of a Markov decision process, and solved via the use of dynamic programming. Unfortunately, computational limits and the curse of dimensionality necessitated the use of a simplified, one-dimensional variation of the task. Despite the simplified nature of this task, the resulting optimal control law exhibited the same qualitative pattern of results as demonstrated by human subjects in the experiment. The comparative results of the minimum jerk model and optimal control model provide further evidence that human performance in the ETA experiment was
characterized by the acquisition of and exploitation of internal models of the cognitive dynamics of the object discrimination task.

7.3 Discussion

7.3.1 Computational models in theory development

A recurring theme of the research conducted in this dissertation is the necessity of combing empirical study, optimality analysis, and cognitive modeling in order to investigate hypotheses regarding the mechanisms underlying human cognition. Experimental studies have long been the foundation of psychology, but the contributions of ideal performer analyses and cognitive models are less widely appreciated. As reviewed in the second chapter of this dissertation, numerous studies have been conducted in recent years on the nature of routine interactive behavior. Yet the interpretations of the findings from these studies have resulted in a wide array of explanations and putative mechanisms. When theories of complex phenomena are relegated to abstract verbal labels, it is often not clear if two competing explanations are in complete agreement or violent contradiction. Computational cognitive models enforce a minimum standard of explicitness in theory development—the theory has to be stated explicitly enough that it can be instantiated in valid computer code. A more interesting use of cognitive models comes from their ability to generate quantitative predictions that can be compared against human performance. The sufficiency of a theory can be objectively determined by whether some cognitive model implementing the theory is actually able to exhibit the behavior in question. Here, the ability to model human learning can provide as valuable constraint on the space of possible cognitive mechanisms as the ability to model asymptotic (expert) human performance. The failure to capture human learning rates may provide valuable insight as to the background knowledge humans might bring in to the experiment that is not captured by the computational model.

Unfortunately, complex models often necessitate the use of free parameters to capture or summarize processes that cannot be directly observed. The resulting danger lies in the misleading assumption that a good fit provides strong evidence, or indeed any evidence, that the mechanisms behind a particular model reflect the reality of the human
cognitive system (Roberts & Pashler, 2000). Consequently, the cognitive models developed in this research have avoided the use of free parameters to the extent possible, and when necessary, treated parameters as predictions that could be tested in further experiments. A free parameter implies that an experimental manipulation that should alter this parameter should result in a predictable change in observed performance. This approach was adopted for one of the bias parameters incorporated in the cognitive model of the first Bayes’ Ball experiment.

However, a more interesting role for cognitive models lies in their ability to explore the consequences of complex ideas (McClelland, in press). If a cognitive model is endowed with the ability to predict the consequences of its actions, how should its control processes be organized in order to make efficient use of this capacity? The adaptive internal model framework was proposed as a plausible answer to this question. The cognitive models developed in this thesis using the adaptive internal model framework were built to perform the task as well as possible, subject to the constraint that they possess only the information given to human subjects, and learn only through experience in the task. By and large, this resulted in an excellent fit to human data.

At the same time, it must be acknowledged that any cognitive model is fundamentally subject to the identifiability problem (J. R. Anderson, 1990), which essentially states that for any behavioral data set there are an infinite number of alternative mechanisms that could reproduce the data equally well. With regard to the potential for cognitive models to definitively uncover the structure or operation of some particular cognitive mechanism, “it is time we stopped fooling ourselves” (J. R. Anderson, 1990; p. 24). One response to the identifiability problem is the incorporation of additional constraints on the space of possible mechanisms, through neuroanatomical and neurocomputational research. To the extent possible, research on the function and anatomy of the cerebellum, cortex, and basal ganglia were offered in support of the adaptive internal model framework developed in this thesis. However, given the current state of progress, it seems unlikely that any theory in cognitive science could be accepted or rejected based solely on known biology.

A more important response to the identifiability problem is the use of principled investigation of optimal task performance, subject to known constraints and limitations
on the cognitive system. In various research traditions, this approach has been referred to as ideal performer or observer analysis, rational analysis, or Bayesian modeling. The goal of each of these is the same. Rather than speculating as to the structure of a cognitive mechanism, one should instead specify the functional properties that this mechanism should possess if it is to optimally accomplish goals in a particular task or task environment. The optimality approach can distinguish limitations in performance from limitations inherent to the information available to the performing system, avoiding needless proliferation of arbitrary mechanisms designed to account for peculiarities of human performance.

The ideal performer approach was used to derive optimality predictions for each of the experiments presented in this dissertation. In the case of the ETA paradigm, computational limitations necessitated deriving optimal performance for a simplified version of the task. Despite the simplification, the resulting optimal control model provided a compelling account for the qualitative features observed in the empirical data.

7.3.2 Implications for other research

The theoretical construct of an internal model has played a central role throughout this dissertation. The particular definition of an internal model used here was borrowed from research on low-level motor control, as in this research tradition internal models are rather precisely defined and have been carefully investigated. On the surface, it might appear strange that internal models have not been given the same emphasis in domains outside of motor control. This discrepancy is perhaps due to the relatively narrow scope of an internal model of one’s motor system—properties such as the mass of a limb, joint configurations, moments of inertia, and other features can be measured in a straightforward manner. By contrast, it is not as simple to precisely define the scope and extent of an internal model of one’s cognitive system, or of the external task environment. The robotics work of Brooks (1991) can be seen as a reaction against the need for such an all-encompassing internal model. In particular, Brooks has argued that, “explicit representations and models of the world simply get in the way. It turns out to be better to use the world as its own best model” (p. 139).
In contrast to this position, this dissertation has demonstrated that the construct of internal models, when narrowly defined, can provide invaluable insight into the nature of the control and coordination processes underlying routine interactive behavior. At the same time, this research has shown that internal models may not be a strictly motor construct, but rather a control construct utilized throughout the cognitive system. This parallels recent work demonstrating that the cerebellum, once thought to be a brain structure dedicated to motor processing, may acquire internal models of a wide range of non-motor entities and play a significant role in all cognitive operations (Bower & Parsons, 2003; Courchesne & Allen, 1997). Researchers of motor control may benefit from the realization that in making reference to internal models, they are no longer studying the motor system narrowly defined, but rather an integrated cognitive-motor system.

At the same time, researchers not specifically interested in human motor control may benefit from the bridges being formed between the seemingly independent domains of cognition, perception, and motor control. In developing their model of simple task switching, Altmann & Gray (2008) found it necessary to posit that the cognitive system can dynamically adjust the activation threshold for retrieving facts from declarative memory based on the features (dynamics) of the ongoing task. Based on the behavioral evidence available, the authors were unable to provide any insight as to the means by which this was accomplished in terms of a computational model. Further, when viewed in isolation, such a capacity might be viewed as an overly complicated mechanism for a relatively narrow phenomenon. Yet when the concept of an internal model is defined not as a motor construct, but rather as a construct that predicts and anticipates ongoing processing—be it cognitive, perceptual, or motor—it is found that the same computational mechanism can provide a parsimonious account for findings in motor studies as well as ‘purely cognitive’ task switching experiments.

Finally, a thorough understanding of the mechanisms underlying the human capacity for fluid and efficient behavior in routine interactive tasks would hold a number of practical and important applications. Motor vehicle collisions and fatalities could potentially be reduced by onboard systems that can predict or ‘trace’ in real time the behavior of human drivers. One such scenario is the incidence of accidents caused by
driving too fast on highway off-ramps, due to applying inaccurate internal models of road conditions and vehicle dynamics. Similarly, research on clinical treatment for disorders of routine interactive behavior, such as “dysmetria of thought” (Gottwald et al., 2004; Schmahmann, 2004) would almost certainly benefit from a more complete understanding of the functional systems involved in non-impaired performance.

7.4 Conclusions

This thesis has argued for a particular theory of the mechanisms underlying the fluent intelligence of human behavior in routine interactive tasks. The central construct in this theory, an internal model, was not invented purely to account for the empirical findings from the experiments reported here. Rather, the construct was borrowed from research on motor control, where it already has gained substantial behavioral and biological support. The experiments conducted in this thesis were designed to test the extent to which this same construct could be recast as a central component in unified theories of embodied cognition.

In addition to the cognitive models developed as formal instantiations of this theory, this research has relied heavily on the use of optimality analysis of human performance. It is worth pointing out that in each case, the derivation of optimal performance required the availability of the mechanism under consideration; namely, a predictive model of embodied task dynamics. In the case of the Bayes’ Ball paradigm, the optimality analysis required the availability of an internal model relating temporal durations to variability in the resulting ball position. In the ETA paradigm, the dynamic programming solution required an explicit probabilistic model of the task dynamics (both motor and cognitive). The fact that human performance can be described as an approximation to the optimal predictions lends support to the hypotheses of this dissertation—that humans possess and exploit such internal models. The convergent use of empirical study, optimality analysis, and cognitive modeling provides substantial evidence in favor of the argument that routine human activity is characterized by the acquisition and exploitation of rich internal models of embodied task dynamics.
References


Appendix A: Debriefing questionnaire used for experiment 1

During the test phase of the experiment the green target was located at one of three distances: near, middle, or far from the ball’s starting position.

1) For **just the near targets**, if you had to estimate, what percentage of trials did you get the ball in the green target zone?

2) Similarly for just the **middle targets**, what percentage of trials did you get the ball in the target zone?

3) How about for the **far targets**? What percentage of trials did you get the ball in the target zone?

4) To estimate the ball’s position, did you consistently use any type of verbal strategy, such as counting out loud or counting silently in your head?

5) To estimate the ball’s position, did you consistently use eye movements across the screen as a simulation of the ball’s position and velocity?

6) **If both**, what determined when you used one strategy or the other, or did you simultaneously use both strategies?

7) Did you use any other type of strategy that wasn’t covered in the previous questions?
Appendix B: Derivation of the ideal performer analysis

This appendix provides the derivation for the probability distribution of the ball’s location \( p(x|\tau) \) starting from two assumed distributions: \( p(x|t) \) relating the ball’s location given the objective elapsed time, and \( p(t|\tau) \), the distribution relating subjective elapsed time \( \tau \) to the amount of physically elapsed time \( t \).

From the rule for marginalizing over a joint probability distribution, it follows that

\[
p(x,\tau) = \int_p(x,t,\tau)dt.
\]

From the chain rule of probability, it is also true that

\[
p(x,\tau) = p(x|\tau)p(\tau).
\]

The above two equations are combined to yield

\[
p(x|\tau) = \int_p(x,t,\tau)dt \quad p(\tau) = \frac{\int_p(x,t,\tau)dt}{p(\tau)}.
\]

By applying the chain rule twice, it is possible to rewrite the joint distribution inside the integral as

\[
p(x,t,\tau) = p(x|\tau)p(t|\tau)p(\tau).
\]

Combining the previous two expressions, the term \( p(\tau) \) cancels in the numerator and denominator, leaving

\[
p(x|\tau) = \int_p(x|\tau)p(t|\tau)dt.
\]

Mathematica version 6.0 (Wolfram Research, 2007) was used to analytically solve the above integral given the two distributions \( p(x|t) \) and \( p(t|\tau) \), yielding the result

\[
p(x|\tau) = \int_0^\infty p(x|t)p(t|\tau)dt
\]
\[
U(\tau) = \int_{\mathcal{A}} u(x) p(x|\tau) \, dx,
\]
where \( u(x) \) is the penalty/reward function defined by the current trial configuration. In the case of three point regions of values \( r_1, r_2 \), and \( r_3 \) with boundaries defined by \((-\infty, x_1, x_2, \infty)\), an analytical solution for the expected utility function can be obtained:

\[
U(\tau) = \frac{1}{2} (r_1 + (r_1 - r_2) \operatorname{Erf}\left(\frac{x_1 - v \tau}{\sqrt{2} (v \sigma_0 + v \sigma_1 \tau)}\right) + r_2 \operatorname{Erf}\left(\frac{x_2 - v \tau}{\sqrt{2} (v \sigma_0 + v \sigma_1 \tau)}\right) + r_3 \operatorname{Erfc}\left(\frac{x_3 - v \tau}{\sqrt{2} (v \sigma_0 + v \sigma_1 \tau)}\right)
\]
where \( \operatorname{Erf} \) and \( \operatorname{Erfc} \) are the Gauss error function and complementary error functions, respectively. The maximum of this expected utility function, corresponding to the decision strategy \( \tau_{opt} \), was obtained numerically for each subject and penalty configuration.
Appendix C: Derivation of the modified minimum jerk model

According to the minimum jerk model (Flash & Hogan, 1985; Hogan, 1984), human reaching movements are planned offline in order to maximize the smoothness of movement, where smoothness is defined as squared jerk (third temporal derivative of hand position), integrated over the course of the entire movement. (Flash & Hogan, 1985) derived a variant of this model applicable to via-point reaching tasks, where the hand must move from a starting position to a final position and pass through a specified intermediate location. However, their derivation assumed that the hand was required to stop at the final target, e.g., zero velocity and acceleration. In the third experiment of this dissertation, subjects were required to execute a similar via-point reaching movement; however, there was no constraint placed on the terminal velocity or acceleration, such that subjects merely needed to pass through the final target. The elimination of this constraint on the motor system adds an additional control dimension that subjects were able to exploit in order to improve their performance. Given the time constraints placed on the movement, performance is defined as being able to move the mouse cursor from the waypoint to the final target more quickly, while maintaining a sufficient level of accuracy.

In this appendix, a modified minimum jerk model is derived that predicts the maximally smooth movement profile for a via-point reaching task, while placing no constraints on terminal velocity or acceleration. This analysis is limited to planar movements. The state of the hand is represented using a time-varying vector of position, velocity, and acceleration of the end-effector in the x and y directions:

\[ s(t) = [x(t), y(t), \dot{x}(t), \dot{y}(t), \ddot{x}(t), \ddot{y}(t)]^T. \]

The state of the system is constrained to evolve according to a known set of dynamics, \[ \dot{s}(t) = f[s(t), u(t), t] \], based on the current state, and a possibly time-varying control signal \( u(t) \). As a rough approximation, it is assumed that humans directly control the change in acceleration (jerk) of the hand through the control signals \( u_x(t) \) and \( u_y(t) \). This results in the following differential equation describing the dynamics of the state

\[ \dot{s}(t) = [\dot{x}(t), \dot{y}(t), \ddot{x}(t), \ddot{y}(t), u_x(t), u_y(t)]^T. \]
At each moment in time, the system is ‘penalized’ using a quadratic cost function of the squared jerk,

\[ L(t) = \frac{1}{2} \left( u_x(t)^2 + u_y(t)^2 \right), \]

and the total cost for a movement is obtained by integrating the cost along the entire path

\[ J = \int_0^T L(t) \, dt. \]

In this application, the cost depends only on the control signals. In a more general setting, the cost can depend on time and the current state as well as final states of the system. The above expression for \( J \) represents a cost functional, which is a mapping from an arbitrary time-varying function to a scalar quantity that is to be minimized. This functional is to be minimized using techniques from the calculus of variations (Bryson & Ho, 1975; Todorov, 2007). Since the state is constrained to obey the system dynamics, it is required that \( \dot{s}(t) - f[s(t), u(t), t] = 0 \). This constraint can be incorporated by adjoining it to the performance index \( J \) using a time-varying co-state vector \( \lambda(t) \). The coefficients of this vector are similar to Lagrange multipliers in constrained optimization. The extended performance index is then given by

\[ \bar{J} = \int_0^T \left[ L(t) + \lambda^T(t) \left( f[s(t), u(t), t] - \dot{s}(t) \right) \right] \, dt. \]

For later convenience, a scalar function \( H \) (known as the Hamiltonian) is defined as

\[ H(t) = L(t) + \lambda^T(t) f[s(t), u(t), t]. \]

The Hamiltonian is substituted into the extended performance index \( \bar{J} \), and the equation is re-arranged using integration by parts to give

\[ \bar{J} = -\lambda^T(t_f)s(t_f) + \lambda^T(t_0)s(t_0) + \int_0^T \left( H(t) + \dot{\lambda}^T(t)s(t) \right) \, dt. \]

For \( \bar{J} \) to be at a minimum, it is necessary that the performance index not decrease for small variations in the control signal (otherwise the new control signal would be more optimal). Stated formally, the first variation in \( \bar{J} \) must equal zero. For the problem type considered here, it can be shown (Bryson & Ho, 1975) that this is achieved when
\[ \dot{\lambda}^T = - \frac{\partial H}{\partial s} \]
\[ \lambda^T (t_f) = 0 \]
\[ \frac{\partial H}{\partial u} = 0. \]

These three equations are known as the Euler-Lagrange equations, and a solution to this set of equations will constitute a control signal and trajectory that achieves a minimum in the cost function. With the initial state of the system specified, this is equivalent to a two-point boundary value problem in differential equations. For a via-point reaching task, an additional constraint that must be satisfied is that the hand position must pass through the waypoint at some unspecified time \( t_1 \). This can be enforced by requiring that \( N(t) = [x(t_1), y(t_1)]^T - [x_1, y_1]^T = 0 \), where \((x_1, y_1)\) specify the coordinates of the waypoint. This additional constraint is adjoined to the cost function using an additional set of multipliers \( \pi = [\pi_1, \pi_2]^T \). This results in a three-point boundary value problem, but also introduces discontinuities in the co-state vector \( \lambda(t) \) and the Hamiltonian. To deal with the discontinuities, the Hamiltonian is defined separately for the initial portion of the movement \( H^-(t), 0 \leq t \leq t_1 \) and the portion after the waypoint \( H^+(t), t_1 < t \leq t_f \). The state and co-state vectors are similarly split into \( S^-(t), S^+(t), \lambda^-(t) \) and \( \lambda^+(t) \). It can be shown that the Hamiltonian and the co-state vectors at time \( t_1 \) must satisfy
\[ \lambda^-(t_1) = \lambda^+(t_1) + \pi^T \frac{\partial N}{\partial S(t_1)} \]
\[ H^-(t_1) = H^+(t_1) - \pi^T \frac{\partial N}{\partial t_1}. \]

In addition, at the waypoint it is required that the state vector is continuous, such that \( S^-(t_1) = S^+(t_1) \). At the final time, the position of the hand is constrained to \((x_f, y_f)\), while the velocity and acceleration are left unspecified. This is achieved by requiring that the co-state multipliers for velocity and acceleration at \( t_f \) equal zero.
Altogether, the preceding derivation specifies a set of differential equations and boundary conditions given by the Euler-Lagrange equations, constraints specified by the initial state, the terminal state, as well as the interior point constraints. This system of equations was solved analytically using the DSolve routine in Mathematica (Wolfram Research, 2007). The resulting horizontal component of the state trajectory minimizing the cost function is given as follows:

\begin{align*}
x^- (t) &= (-24 t_i^3 (t_i^2 - 5 t_i t_f + 10 t_f^2) x_o + t^5 (x_o - x_i) + \\
&10 t_i^3 x_o - x_i) + 5 t^4 t_i (-x_o + x_i) + \\
& t^5 (t - t_i) t_i^3 (t_i - t_f) (2 t_i (t_i - 4 t_f) + t (t_i + 2 t_f)) / (24 t_i^3 (t_i^2 - 5 t_i t_f + 10 t_f^2)) \\

\dot{x}^- (t) &= (t^2 (-120 (t^2 - 4 t t_f + 6 t_f^2)) (x_o - x_i) + \\
& t_i^3 (t_i - t_f) (4 t t_i (t_i - 10 t_f) - 6 t_i^2 (t_i - 4 t_f)) + \\
& 5 t^2 (t_i + 2 t_f)) / (24 t_i^3 (t_i^2 - 5 t_i t_f + 10 t_f^2)) \\

\ddot{x}^- (t) &= (-120 t (t^2 - 3 t t_f + 3 t_f^2) (x_o - x_i) + t t_i^3 (t_i - t_f) \times \\
& (3 t_i (t_i - 10 t_f) - 3 t_i^2 (t_i - 4 t_f) + 5 t^2 (t_i + 2 t_f)) / (6 t_i^3 (t_i^2 - 5 t_i t_f + 10 t_f^2)) \\

x^+ (t) &= (120 (-t^5 - 5 t^4 t_f + 10 t^3 t_f^2 - 6 t_f^5) (x_o - x_i) + \\
&t_i^5 x_o - 5 t_i^3 (t_i - 2 t_f) x_f + t_i^5 (t - t_f) (6 t_i^4 - 10 t_i^3 + \\
& 5 t_i^2 + (-24 t^3 + 50 t t_i^2 - 35 t_i^3) t_f + 4 (9 t^2 - 25 t t_i + \\
& 25 t_i^2) t_f^2 + 4 (9 t_f - 25 t) t_f^3 + 36 t_f^4) / (24 t_i^3 (t_i^2 - 5 t_i t_f + 10 t_f^2)) \\

\dot{x}^+ (t) &= (-120 t^2 (t^2 - 4 t t_f + 6 t_f^2) (x_o - x_i) + \\
&t_i^5 (6 t^4 - 24 t^3 t_f + 36 t^2 t_f^2 - \\
& 4 t t_i (t_i^2 - 5 t_i t_f + 10 t_f^2) + t_i^2 (t_i^2 - 5 t_i t_f + 10 t_f^2)) / (24 t_i^3 (t_i^2 - 5 t_i t_f + 10 t_f^2)) \\

\ddot{x}^+ (t) &= -(120 t (t^2 - 3 t t_f + 3 t_f^2) (x_o - x_i) + \\
&t_i^5 (-6 t^4 + t_i^3 + 18 t^2 t_f - 5 t_i (t_i - 2 t_f) t_f - \\
& 18 t t_f^2) / (6 t_i^3 (t_i^2 - 5 t_i t_f + 10 t_f^2)) 
\end{align*}
Equivalent expressions for the $y$ coordinate of the trajectory are obtained by replacing all references to $x$ in the above equations by their $y$ equivalents, and replacing references to $\pi_1$ with $\pi_2$. The values of the constants $\pi_1$ and $\pi_2$ are obtained by analytically solving the continuity requirement

\[
x^-(t_1) = x^+(t_1) = x_i,
\]
\[
y^-(t_1) = y^+(t_1) = y_i.
\]

Solving the above yields the solution

\[
\pi_1 = (120 \ (6 \ t_f \ (-x_0 + x_i) + t_1^5 (x_0 - x_f) + 10 \ t_1^3 t_f^2 (x_0 - x_f) + 5 \ t_1^4 t_f (-x_0 + x_f))/ (t_1^5 (t_1 - t_f)^2 (t_1^3 - 8 \ t_1^2 t_f + 28 \ t_1 \ t_f^2 - 36 \ t_f^3))
\]
\[
\pi_2 = (120 \ (6 \ t_f \ (-y_0 + y_i) + t_1^5 (y_0 - y_f) + 10 \ t_1^3 t_f^2 (y_0 - y_f) + 5 \ t_1^4 t_f (-y_0 + y_f))/ (t_1^5 (t_1 - t_f)^2 (t_1^3 - 8 \ t_1^2 t_f + 28 \ t_1 \ t_f^2 - 36 \ t_f^3)).
\]

Finally, it remains to be seen how the passage time through the waypoint, $t_1$, is obtained. For a given movement task, consisting of a specified initial state, target and waypoint information, the optimal value for $t_1$ can be obtained numerically by finding the roots of the fifth-order polynomial given by

\[
\pi_1 \dot{x}^-(t_1) + \pi_2 \dot{y}^-(t_1) = 0,
\]

and accepting only the real roots where $0 < t_1 < t_f$. A Mathematica function for numerically computing this value is provided in Table C1.
Table C1. Mathematica function for computing the optimal waypoint passage time for the modified minimum jerk model of motor planning.

```mathematica
WaypointPassageTime[x0_, y0_, x1_, y1_, xf_, yf_, tf_] := Module[
    {pi1, pi2, vx, vy, t1},

    pi1 = (120 (6 tf^5 (-x0 + x1) + t1^5 (x0 - xf) +
        10 t1^3 tf^2 (x0 - xf) + 5 t1^4 tf (-x0 + xf)))/(t1^5 (t1 -
        tf)^2 (t1^3 - 8 t1^2 tf + 28 t1 tf^2 - 36 tf^3));
    pi2 = (120 (6 tf^5 (-y0 + y1) + t1^5 (y0 - yf) +
        10 t1^3 tf^2 (y0 - yf) + 5 t1^4 tf (-y0 + yf)))/(t1^5 (t1 -
        tf)^2 (t1^3 - 8 t1^2 tf + 28 t1 tf^2 - 36 tf^3));

    vx[t_] := (t^2 (-120 (t^2 - 4 t tf + 6 tf^2) (x0 - x1) +
        t1^3 (t1 - tf) (4 t t1 (t1 - 10 tf) - 6 t1^2 (t1 - 4 tf) +
        5 t^2 (t1 + 2 tf)) pi1))/(24 t1^3 (t1^2 - 5 t1 tf +
        10 tf^2));
    vy[t_] := (t^2 (-120 (t^2 - 4 t tf + 6 tf^2) (y0 - y1) +
        t1^3 (t1 - tf) (4 t t1 (t1 - 10 tf) - 6 t1^2 (t1 - 4 tf) +
        5 t^2 (t1 + 2 tf)) pi2))/(24 t1^3 (t1^2 - 5 t1 tf +
        10 tf^2));

    t1 /. FindRoot[pi1 vx[t1] + pi2 vy[t1] == 0, {t1, 0.5 tf}]
]
```

Appendix D: Electronic archive of materials

An electronic archive of all materials used or generated during the course of this research is available by contacting the author, Chris R. Sims, via email at simsc@rpi.edu. In addition, the archive is also available on a server maintained by the CogWorks Laboratory at Rensselaer Polytechnic Institute. For contact information, please go to http://www.cogsci.rpi.edu/cogworks/.

The electronic archive contains the experimental software and human subject materials (task instructions, survey and consent forms) for all experiments reported in this dissertation, the empirical results and statistical analyses, as well as all code used for the cognitive models and mathematical optimality analyses.