A Model of Dynamic Choice, Confidence, and Motor Response

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Kenneth C. Olson, B.S., M.A.

Graduate Program in Psychology

The Ohio State University

2011

Dissertation Committee:

Trisha Van Zandt, Advisor
Richard Jagacinski
Thomas Nygren
© Copyright by

Kenneth C. Olson

2011
Abstract

In psychology research, mathematical models have developed to explain choices and response times. More recently, researchers have begun creating models to try explaining confidence as well (e.g., Vickers, 1979; Pleskac & Busemeyer, 2010; Ratcliff & Starns, 2009; Van Zandt, 2000). I join this effort by adapting the Linear Ballistic Accumulator model (Brown & Heathcote, 2008) to account for confidence by employing the relative balance-of-evidence hypothesis (Merkle & Van Zandt, 2006) in my Ballistic Accumulator for Confidence (BAC).

The goal of this research is to account for multiple components of a decision task, including choice probabilities, confidence ratings, and response times. Often in a decision experiment, researchers count the time between stimulus appearance and the press of a button for a choice simply as response time, although many different mental activities occur during that time. Their primary focus is the decision process, but other processes include movement and movement programming.

Through a within-subjects experiment designed to extract movement times from response times to obtain reaction times, I show that decision times and movement times are not independent in determining response times. In some conditions, movement times vary systematically and correlate with reaction times, while in other conditions, they do not.
The dynamic choice models I review assume that all non-decision processes are independent of the decision process and simply add a constant non-decision component to decision time for estimating response time. I found that this assumption is incorrect and that models are impaired by failure to account for movement times. Aimed movement toward a target is explained by several models, the most famous of which are Fitts’ Law (Fitts, 1954) and the optimized-submovement model (OSM) (Meyer et al., 1990). By evaluating a series of related dynamic confidence models, I show that the best model is one which incorporates the OSM. In the model, movement programming is assumed to be part of the decision task and depends on movement times, while movement times depend on the apparatus used for choices and confidence ratings.
Acknowledgments

This work was made possible by NSF grants SES-0437251, BCS-0738059, and SES-1024709. Such financial support was transferred through my mentor, Trisha Van Zandt, whose social and intellectual support solidified my proposed work. Her patience allowed me to explore several directions for my research and my career. Together with Rich Jagacinski and Jay Myung, she often provided advice to reshape this project.

Rich Jagacinski was an important resource for understanding the human factors literature that was a pillar of my study. Furthermore, he and Tom Nygren were very accommodating of the unusual schedule for fulfillment of the graduate school requirements.

I further acknowledge the times that my colleagues, Kevin Potter and Brandon Turner, graciously listened to the ideas presented herein while they were still amorphously composed. I also thank Lindsey Gilmore, Justin Braun, Katherine Sasso, Sonia Saberi, Sasha Saberi, Mary Reamsnyder, and Samuel Rowe for their assistance with a pilot study.

I cannot leave out my parents, Carl and Sherri, who have been understanding always of me, although not always of my work. They morally backed my success even as it reduced my contact with them.
More than anyone, my wife should be credited for helping me in this endeavor. Nicole postponed some of her own pursuits to hasten mine. Her stable presence promoted my mental health, her generous efforts in nonacademic activities maintained my pace of work, and her appreciation for cerebral preoccupation encouraged my completion of this research.
Vita

2004 ........................................... B.S. Economics & Psychology,
Baker University, Kansas
2008 ........................................... M.A. Psychology,
The Ohio State University
2005-2006 ................................. University Graduate Fellow,
The Ohio State University
2006-2011 ................................. Graduate Teaching Associate,
The Ohio State University
Summers, 2007-2010 ........................ Graduate Research Assistant,
The Ohio State University

Fields of Study

Major Field: Psychology

Minor Field: Cognitive Modeling
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iv</td>
</tr>
<tr>
<td>Vita</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>x</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xi</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Response Times</td>
<td>5</td>
</tr>
<tr>
<td>1.1.1 Accumulator Model</td>
<td>8</td>
</tr>
<tr>
<td>1.1.2 Poisson Race Model</td>
<td>9</td>
</tr>
<tr>
<td>1.1.3 Diffusion Model</td>
<td>10</td>
</tr>
<tr>
<td>1.1.4 Linear Ballistic Accumulator Model</td>
<td>12</td>
</tr>
<tr>
<td>1.2 Confidence</td>
<td>14</td>
</tr>
<tr>
<td>1.2.1 Expanded Poisson Race Model</td>
<td>19</td>
</tr>
<tr>
<td>1.2.2 RTCON Model</td>
<td>20</td>
</tr>
<tr>
<td>1.2.3 2DSD Model</td>
<td>22</td>
</tr>
<tr>
<td>1.2.4 Modified Linear Ballistic Accumulator Model</td>
<td>23</td>
</tr>
<tr>
<td>1.3 Non-Decision Times</td>
<td>24</td>
</tr>
<tr>
<td>1.3.1 Movement Times</td>
<td>26</td>
</tr>
<tr>
<td>1.3.2 Movement Programming Times</td>
<td>34</td>
</tr>
<tr>
<td>1.4 Goals of the Research</td>
<td>39</td>
</tr>
<tr>
<td>1.5 Organization of this Dissertation</td>
<td>42</td>
</tr>
</tbody>
</table>
5.4 Conclusion ....................................................... 128

Appendices .................................................. 130

A. Appendix .................................................. 130
   A.1 Formulas ............................................. 130
   A.2 Instructions to Participants for Targeting Conditions ........ 134
   A.3 Instructions to Participants for Confidence Conditions .... 135
   A.4 Model Fits .......................................... 137

References ................................................... 142
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Table of parameters</td>
</tr>
<tr>
<td>2.2</td>
<td>Other nomenclature</td>
</tr>
<tr>
<td>3.1</td>
<td>Latin square for the ordering of conditions</td>
</tr>
<tr>
<td>4.1</td>
<td>Model fits</td>
</tr>
<tr>
<td>A.1</td>
<td>Parameter estimates from model fits</td>
</tr>
<tr>
<td>A.2</td>
<td>Parameter estimates from MT model fits</td>
</tr>
</tbody>
</table>
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Times for components of the decision task</td>
<td>6</td>
</tr>
<tr>
<td>2.1 Modified LBA</td>
<td>49</td>
</tr>
<tr>
<td>2.2 Drift rate distribution</td>
<td>50</td>
</tr>
<tr>
<td>3.1 Distributions of stimuli</td>
<td>65</td>
</tr>
<tr>
<td>3.2 Ideal accuracy</td>
<td>67</td>
</tr>
<tr>
<td>3.3 Sequence of events for a trial in rectangular-regions-for-targeting condition</td>
<td>70</td>
</tr>
<tr>
<td>3.4 Sequence of events for a trial in triangular-regions-for-confidence condition</td>
<td>71</td>
</tr>
<tr>
<td>3.5 Distribution of target levels</td>
<td>77</td>
</tr>
<tr>
<td>3.6 Accuracy over trials</td>
<td>80</td>
</tr>
<tr>
<td>3.7 Examples of participants' confidence distributions</td>
<td>82</td>
</tr>
<tr>
<td>3.8 Reduced response time distributions</td>
<td>84</td>
</tr>
<tr>
<td>3.9 Three decision measurements for the range of stimuli</td>
<td>85</td>
</tr>
<tr>
<td>3.10 RT quantiles for levels of confidence/targeting</td>
<td>88</td>
</tr>
<tr>
<td>3.11 MT quantiles for levels of confidence/targeting</td>
<td>90</td>
</tr>
</tbody>
</table>
3.12 Predicted results for MT and RT ........................................ 94
3.13 Empirical results for MT and RT ...................................... 95

4.1 Predictions from Fitts’ Law of MT quantiles for levels of confidence/targeting based on nominal index of difficulty ............................... 103
4.2 Predictions from Fitts’ Law of MT quantiles for levels of confidence/targeting based on effective index of difficulty ............................... 104
4.3 Predictions from the OSM of MT quantiles for levels of confidence/targeting based on nominal index of difficulty ............................... 105
4.4 Predictions from the OSM of MT quantiles for levels of confidence/targeting based on effective index of difficulty ............................... 106
4.5 Predictions from BasicRespR of response time quantiles for levels of confidence/targeting .................................................... 111
4.6 Predictions from BasicRespT of response time quantiles for levels of confidence/targeting .................................................... 112
4.7 Predictions from BasicReacR of RT quantiles for levels of confidence/targeting .................................................... 113
4.8 Predictions from FProgReacR of RT quantiles for levels of confidence/targeting .................................................... 114
4.9 Predictions from CProgReacR of RT quantiles for levels of confidence/targeting .................................................... 115
4.10 Predictions from CProgReacT of RT quantiles for levels of confidence/targeting .................................................... 116
Chapter 1: INTRODUCTION

A substantive issue in the study of human thought and behavior is the nature of decision making. In mathematical psychology, which often studies the same content as cognitive science, theories for decision making take the form of mathematical models. Because the inner workings of the mind are hidden from direct view, those observable actions associated with decision making are related to each other by a suggested computation. As a model, such a computation is a simple substitute for an unseen process.

The genesis of many mathematical models of judgment and decision making is psychophysics, where Gustav Fechner was a pioneer in quantifying psychological variables (Coombs, Dawes, & Tversky, 1970). For example, Thurstone’s model of comparative judgment followed psychophysical methods in the 1920s by evaluating data from human subjects who had been asked to repeatedly gauge the difference in some physical property between paired stimuli. From these trials, experimenters could construct psychometric functions showing the probability of choosing one response over another, given the difference between stimuli, which generally took the form of an S-shaped or sigmoid curve (Vickers, 1979). The choice was viewed as a problem of discrimination between values associated with stimuli, where larger distances between values on the scale make discrimination between stimuli easier.
The most important contribution of Thurstone’s model (1927) was in explaining how errors arise in human judgment by assuming that upon repeated, independent presentations, each stimulus does not give rise to identical magnitudes of subjective experience; rather, subjective sensations regarding a stimulus follow a normal unidimensional distribution. Some other notable theories of decision making address the core decision component with additional examination of sensation. For example, Duncan Luce’s constant utility model (1959) and signal detection theory (Tanner & Swets, 1954) had a lasting influence on decision-making research by accounting for empirical patterns of choice probabilities, but they said little about the real-time decision process or its related effects.

The whole human decision process spans from a person’s attention to a problem to completion of an action, but psychologists have often explored each component of the process separately. They may study components such as perception, memorization, response mapping, or motor responding, but rarely in unison. However, some efforts have been made to develop a theory to explain how more than two pieces fit together. Because my purpose here is to account for several components of the decision process in a theory of choice, I will review some of these previous efforts below.

The primary measurements of decisions are the frequencies of particular responses, which allow for estimation of the probability of each choice. Combining these with secondary measurements allows researchers to infer more about what occurs during the decision process. Two common secondary measurements are confidence judgment and response time. Two others among these secondary measurements which I focus this research on are movement time and movement programming time. The main aim of my research is to not only account for choice probability and confidence in a decision
task, but also to dissect response time into decision processing time, movement time, and movement programming time (See Figure 1.1).

The work I present is preceded by the work of eminent researchers such as Joseph Jastrow in the nineteenth century (Peirce & Jastrow, 1884), who investigated reaction time (RT), the time between the presentation of a stimulus and beginning of a basic response to it. He saw RTs as measures of individual differences in humans’ mental activities.

Franciscus Donders, an ophthalmologist credited as the father of “mental chronometry”, was the first to use differences in human response time to infer differences in cognitive processing (Donders, 1868). Donders used three primary tasks in his experiments: simple reaction, discrimination, and choice. In a simple reaction task, a person makes a response as soon as a stimulus is presented. In a discrimination task, a person makes a response to one particular stimulus but not to other stimuli that are presented. In a choice task, a person makes a different response to each type of stimulus presented. Donders developed the subtraction method (1868) to analyze the mental processing components of human performance of these tasks.

The subtraction method assumes that different tasks share many components but are differentiated by the components that are subtracted from them. The method further assumes that the time to complete each component is independent of the other components. For example, if simple reaction and choice processes are identical except for the insertion of “discrimination” and “selection” operations in the choice task, then choices will take longer than simple reactions only for the time added from those extra operations. By subtracting the time to simply react to a stimulus from
the time to make a choice about a stimulus, Donders inferred the time needed to complete discrimination and selection operations for the choice.

By Donders’s theory, the time it takes to perform a task depends on the number and types of mental components involved. The assumption of ”pure insertion”, that the times for processing components are independent of each other so that one can be inserted into a task without changing the others, is not fully supported. Miller and Low (2001) showed that for the classic tasks Donders used, the times to complete motor processing stages are roughly independent of the tasks being completed, but for more complicated tasks they may not be independent. Donders’s choice tasks did not involve the secondary measurement of confidence. As I will discuss later, typical confidence tasks have more elaborate movement requirements than Donders’s tasks, which are not independent of the confidence responses.

In my research I am concerned with how movement time influences response time and how both are influenced by confidence. By the subtraction method, one could measure the time to finish the movement that indicates a confidence rating and the time to begin that same movement, subtract the latter from the former, and assume that the difference has no influence on the subtrahend. In other words, movement time has no influence on the total response time minus movement time (reaction time). I plan to test whether this is true.

The problem this paper addresses is how to insulate response time from the influence of movement when multiple responses are possible. Assuming that movement has a strong impact on response times even when movement times are subtracted, I propose a way to eliminate that impact experimentally and a way to account for
that impact in dynamic choice models. The problem is especially pertinent to con-
fidence judgments because the number of response alternatives is greater than it is
for choices. Confidence complicates the response apparatus that must be created and
has the potential to increase response times and movement times — and even entan-
gle them. The difficulties are especially obvious for graphic rating scales for which
confidence is measured as a distance.

The first objective is to build my own model to account for choices, confidence
judgments, and response times. From there I can augment the model to account
for movement times. My second objective is to show by fitting different versions of
the model to data from a decision task that a model with the added structure for
movement times of responses is superior to a model without the added structure. My
third objective is to reduce or eliminate the influence of movement on response times
through experimental manipulations.

The present review of psychology literature consists of sections primarily grouped
by models of choice response time, models of confidence, and models of movement
time. The experiment I’ve designed and model I’ve developed combine ideas from
each section with special attention to how movement time interacts with response
time.

1.1 Response Times

In this section, I review four models that explain the regularities in choice response
times: Vickers’s accumulator model, the Poisson race model, the diffusion model, and
the linear ballistic accumulator.
Donders was among the earliest researchers to believe that response times can reveal something about underlying mental processes, but later psychologists incorporated response times into a wider range of research (Townsend & Ashby, 1983). Use of response times was mostly limited to specialists in psychophysics through the 1940s, but experimenters like Henmon (1911) had noticed regularities in the response times of human choice decades before.
A very old observation on the relation between response time and choice is called the “speed-accuracy trade-off” (Vickers, 1979). Response time and probability of a correct response share a positive relation (Garrett, 1922; Johnson, 1939). An intuitive theory for why the proportion of correct choices decreases when response time decreases is that less caution on the part of the decision maker causes both changes in variables. If one wants to choose more quickly, then one must accept greater risk of making a mistake; likewise, if one wants to be more certain of making a correct choice, then one must sacrifice some speed.

Another behavioral phenomenon documented for response time and choice is commonly referred to as “slow errors” (Swensson, 1972; Luce, 1986). Their occurrence is very dependent on the choice task, but generally, when accuracy is emphasized over speed, decision times tend to be greater for incorrect choices than for correct choices. Slow errors are not typically expected when speed is emphasized over accuracy (Townsend & Ashby, 1983).

Duncan Luce wrote, “. . . we surely do not understand a choice process very thoroughly until we can account for the time required for it to be carried out” (Luce, 1986, p. vii). A model that can explain both choice probabilities and response times, often called a “dynamic choice” model, is more likely to approximate the choice process. Nevertheless, few models of choice which incorporated response time emerged prior to the late 1960s (e.g., Hick, 1952; Hyman, 1953). Luce’s early theories modeled only final choice states, while his later work considered the response times of choices. Recent dynamic choice models typically have the goal of explaining both the speed-accuracy trade-off and slow errors.
1.1.1 Accumulator Model

Among the most successful dynamic models of choice are the “sequential sampling” models (Ratcliff & Smith, 2004). Sequential sampling models naturally explain the speed-accuracy trade-off in decision making and have become some of the most important of models in cognitive psychology.

Part of a larger order of models called stochastic models, sequential sampling models assume that a person obtains a series of momentary observations or samples of stimuli to make a decision. As these observations occur over time, they are perturbed by random noise. Vickers provides an overview and analysis of several of the models preceding the accumulator model in his book on visual perception (Vickers, 1979). These include the “runs” model (Audley, 1960) and the simple “random walk” model (Laming, 1968). The accumulator model that Vickers presents is the successor to LaBerge’s recruitment model (LaBerge, 1962), in which “evidence” is sampled in favor of competing alternatives. Each alternative has a counter to store the evidence that supports the choice of that alternative. A person decides that stimulus $k$ is bigger than stimulus $j$ if evidence in the $k$ counter reaches its threshold first. When the evidence crosses a counter’s boundary, it determines both the choice and the choice RT.

LaBerge’s recruitment model samples evidence in discrete increments at discrete time steps, which results in poor predictions of response times. Vickers’s accumulator model is adapted to accrue evidence from a continuous, normal distribution but still in discrete time steps (Vickers, 1979). The mean of the evidence distribution for a counter is based on the quality of information from its related stimulus. The mean increases as one stimulus becomes easier to discriminate from another stimulus. A
The shape of the distribution of the response times for choices is generally different for different pairs of stimuli and for correct and incorrect responses (Vickers, 1979). One of the major contributions of the accumulator model is its ability to simultaneously model the empirical relative frequencies of choices, and their response time distributions. Predicted choice probabilities still follow an S-shaped psychometric function while the model places the means of the response time distributions in the proper order. It also roughly describes the skew of those distributions and successfully predicts that when the difference between stimuli is smallest, response times will rise to a maximum.

1.1.2 Poisson Race Model

The Poisson race model (Audley & Pike, 1965; Pike, 1973; Townsend & Ashby, 1983; Van Zandt, 2000; Van Zandt, Colonius, & Proctor, 2000) provides a different description of the time course of decisions. Unlike the accumulator model, its process accrues discrete amounts of evidence on a continuous time scale. The model assumes that when a person must choose between two options, evidence in support of response A, \( R_A \), increments counter A. Similarly, counter B records evidence in support of response B, \( R_B \). Evidence arrives in bursts like the firing of a neuron and at exponentially distributed time intervals. The exponential distribution of inter-arrival times
means that each counter can be described by a Poisson process. The rates at which counters gather evidence, $v_i$, where $i = A, B$, may vary between counters. When option A is best, the accrual rate for counter $A$ should be greater than the accrual rate for counter $B$, $v_A > v_B$. For easier decisions, the difference between the rates should be greater.

As the person accrues evidence, the two counters race toward thresholds, $K_i$. The counters work simultaneously and independently. Because these are two uncorrelated Poisson processes, the probability that a piece of evidence supports option A is simply $v_A/(v_A + v_B)$. Likewise, the probability that a piece of evidence supports option B is $v_B/(v_A + v_B)$. Given a particular stimulus, $S_j$, we can easily compute the probability that $K_A$ bursts of evidence arrive on counter $A$ while fewer than $K_B$ burst of evidence arrive on counter $B$ by time $t$, which provides the probability of response $A$ at time $t$. To find the overall probability of response $A$, $p(R_A|S_j)$, we have to integrate over time.

A decision time can be found by the sum of the inter-arrival times of evidence before either counter reaches its threshold, the minimum from two gamma distributions.

Through use of thresholds, the model explains the speed-accuracy trade-off. High thresholds produce longer decision times but greater accuracy in choices because more evidence is needed to make a decision. The evidence must pile higher, so the wrong counter is less likely to accidentally finish first, but the right counter also takes longer to finish.

### 1.1.3 Diffusion Model

A distinction can be made among sequential sampling models by the stopping rule each one uses. The accumulator model and Poisson race model have what are
called “absolute” stopping rules so that a decision is made when evidence reaches any criterion for a response. An increase in the amount of evidence for one response does not change the amount of evidence for another response. Random walk models and diffusion models have a “relative” stopping rule so that a decision is only made when evidence for one response exceeds evidence for the other responses by a criterion amount (Ratcliff & Smith, 2004).

Diffusion models can be considered a special case of random walk models (Pleskac & Busemeyer, 2010) in which time steps for evidence sampling are arbitrarily small. Infinitesimal steps help improve the fit of the model over the accumulator model. Ratcliff originally proposed a diffusion model to describe the cognitive process for memory tasks (Ratcliff, 1978) in which evidence comes from past experiences rather than just current sensation. While the accumulator and Poisson models require a separate counter of evidence for each alternative, each with its own threshold, the diffusion model uses a single counter to represent the relation between evidence levels for the alternatives. In a two-choice model, evidence that supports one alternative also contradicts the other. Evidence sampled from a normal distribution with a mean, called the drift rate, $\xi$, and standard deviation, called the diffusion coefficient, $s$, moves the process along a noisy path between the thresholds. The diffusion drift rate is similar to the Poisson evidence accrual rate, but evidence is on a continuous scale.

Like the Poisson race model, the diffusion model accounts for the speed-accuracy trade-off with changes in decision thresholds. In speed conditions, human subjects move their decision criteria to require less evidence for a response, and in accuracy conditions, they move their criteria to require more evidence for a response. Typically,
accuracy instructions cause accuracy to increase by a small fraction and response times to increase by several hundred milliseconds (Ratcliff & Smith, 2004).

Just as the accumulator model included more sources of variability over time (Smith & Vickers, 1988), the diffusion model evolved to have several more parameters to represent variability, particularly the range of the starting point for the process, $s_z$, the range of the non-decision component of response time, $s_t$, and the standard deviation of the drift rate across trials, $\eta$. The many combinations of sources of variability produce a variety of predictions for response time distributions (Ratcliff & Smith, 2004). The model explains slow errors by drift rate variability and the stochastic nature of the process: a drift rate that just happens to be small allows evidence to randomly reach the incorrect threshold first. Setting the parameter this way reflects a difficult decision when the quality of evidence is poor so that people struggle to choose and may do so erroneously. Work on the diffusion model by (Ratcliff & Rouder, 1998) has shown how the addition of parameters $s_z$ and $\eta$ can accommodate errors faster than correct responses in some circumstances and errors slower than correct responses in others. The parameter $s_t$ improves the general shape of the response time distributions.

### 1.1.4 Linear Ballistic Accumulator Model

The dynamic models of choice show a great deal of mimicry (Link, 1992; Luce, 1986; Smith & Vickers, 1988; Townsend & Ashby, 1983; Van Zandt et al., 2000). Early sampling models were fit to mean response times for different choices and conditions, but this hid some of their predictions important for model selection (Luce, 1986). In contrast, the diffusion model explains how the relative speeds of correct and incorrect
decisions change as a function of experimental conditions and why response time distributions take particular shapes. Although the diffusion model evolved quickly to fit response time distributions more precisely, the accumulator and Poisson models have remained competitive. For example, when the many variability parameters of the diffusion model were added to the accumulator model, both performed well (Ratcliff & Smith, 2004).

The increasing complexity of the dynamic choice models must be justified at each step. Additional complexity can provide better model fits at the loss of generalizability (see Myung, 2000), so there is reason to be cautious about each model parameter. In an effort to simplify the leading models of choice and response time, Brown and Heathcote (2005) proposed the ballistic accumulator model. The model appears in many ways to be a special accumulator model, but it is not even in the same class. The ballistic accumulator dispenses with sequential sampling, which gives the sequential sampling processes their psychological meaning. Although other sources of variability are modeled, the deterministic accumulation of information in the ballistic accumulator ignores within-decision variability. Initially, the model had “leakage” of evidence and special response competition, but these properties were later shown to be unnecessary for capturing essential empirical response time phenomena (Brown & Heathcote, 2008). Without the nonlinear components of leaky accumulation of evidence and response competition, the model becomes the linear ballistic accumulator (LBA).

The LBA produces the correct shapes of response time distributions without the assumption that evidence varies randomly from moment to moment. Although other models have used independent, linear accumulators without adequately describing
response time distributions (particularly for incorrect responses) (e.g., Smith & Vickers, 1988; Audley & Pike, 1965; Townsend & Ashby, 1983; Grice, 1972), the LBA can account for the same phenomena as the revised diffusion model (Brown & Heathcote, 2008). The key difference between the LBA and other models that lack the typical stochastic component (e.g., Grice, 1972) is in the accumulation process: other simplified models have been based on a random walk, but the LBA is based on multiple separate accumulators.

In the LBA, $N$ accumulators begin with a starting level of evidence, $a$. Each accumulator has a drift rate, $d$, that pushes evidence toward a threshold, $b$. The first accumulator with evidence above its threshold produces a response. The time taken for an accumulator to reach threshold is $(b-a)/d$. There are two sources of variability from trial to trial: a uniform distribution of starting points and a normal distribution of drift rates. Starting evidence is on the interval $[0, A]$, and drift rates have standard deviation $s$ and means $v_i$, where $i = 1, 2, \ldots N$.

Although the LBA and several sequential sampling models have been successful in accounting for choice patterns and response latencies from empirical studies, they have been long in their development. In the decades of research to model choice response times, confidence in choices has typically been a separate consideration. Confidence has been mathematically modeled, but rarely in conjunction with response times.

1.2 Confidence

The sequential sampling models were introduced to account for choices and response times but have been slow to account for confidence judgments (Vickers, 1979). Vicker’s accumulator model, the Poisson race model, the RTCON model (Ratcliff &
Starns, 2009), and the 2DSD model (Pleskac & Busemeyer, 2010) have done well to address all three measures. These few models of choice, response time, and confidence occupy a unique and valuable space in cognitive science. Although the ambition behind them positions them to develop a deeper understanding of human judgment and decision processes, they also have more ecological validity by accounting for multiple variables. Neglecting confidence in models of choice is a mistake when people commonly follow a decision with a confidence judgment.

Very important judgments almost always have an explicit confidence statement attached to them, such as an eyewitness’s confidence in identification of a criminal (Wells, Ferguson, & Lindsay, 1981; Brewer, Keast, & Rishworth, 2002) and a physician’s certainty that a medical diagnosis is correct (Arkes et al., 1995; Goldberg, 1968).

How a decision process gives rise to feelings of confidence is of great significance to criminal justice. Consider how witnesses to a crime are asked to choose whom they saw commit the act and tell how sure they are of what they remember. If they testify in a later trial, judges and juries will place a great deal of trust in the witnesses’ statements of confidence even when it does not match accuracy of judgment (Wells et al., 1981). Surveys of the legal system show that police officers, attorneys, and jurors hold a common belief that eyewitness confidence is indicative of accuracy in memory (Deffenbacher & Loftus, 1982; Potter & Brewer, 1999).

In medical matters, a job of the general practitioner is to decide what condition afflicts a patient and offer treatment based on symptoms. Some popular entertainment has been based around cases of people with strange health problems who must seek medical assistance. Though the doctor’s role is dramatized, it is not wrong to
emphasize the uncertainty associated with deciding what disease is present and predicting treatment outcomes. “Real-world” studies of confidence judgments include assessment of danger posed by mental health patients (Dawson et al., 1993), prognosis of heart catheterization (Goldberg, 1968), and diagnosis of ulcers (Goldman et al., 1983).

Researchers have long been concerned with how people assess their confidence that choices are accurate or expectations are correct (e.g., Einhorn & Hogarth, 1978; Keren, 1991). Confidence is often treated as a subjective probability estimate that does not always match objective probability (e.g., Dawes, 1980; Griffin & Tversky, 1992; Klayman, Soll, Gonzalez-Vallejo, & Barlas, 1999; Lichtenstein & Fischhoff, 1977), but confidence is not just used to study the correspondence between subjective experience and reality. Like RT, it is expected to reveal characteristics of a decision process. For example, cognitive psychologists sometimes rely on confidence ratings to evaluate theories of memory (e.g., Ratcliff, Sheu, & Gronlund, 1992; Wixted, 2007; Yonelinas, 1994).

The foundation of many models of choice for recognition memory is signal detection theory (SDT). SDT was first applied to recognition memory by Egan (Egan, 1958; Egan, Schulman, & Greenberg, 1959), but the theory was developed earlier. Tanner and Swets (1954) reinterpreted ideas from electrical engineering regarding the detection of when a signal is present in a system. Their theory is the basis for statistical decision-making in which the value of an observation is random, and a conclusion about the origin of the observation comes from its relation to a decision criterion. (A full account of SDT can be found in Macmillan, 2002.)
The global matching models of memory, such as MINERVA2 (Hintzman, 1984), SAM (Raaijmakers & Shiffrin, 1981), and TODAM (Murdock, 1982) contain assumptions about noise in memory. Through SDT, they suggest that information retrieved from memory has variable memory strength and that studied and unstudied stimuli both have a normal distribution of strength with some mean and variance. To decide whether a stimulus has been previously studied, a person sets a criterion. Stimuli with memory strengths greater than the criterion are judged to be previously studied, and stimuli with memory strengths less than the criterion are judged to be new. When applied to psychology, SDT makes explicit in its parameters the distinction between sensation components and response components of choices (Coombs et al., 1970). It claims that no hard threshold for sensation exists and instead criteria for responding vary depending on circumstances.

For studies of recognition memory, confidence ratings may be integrated with choice in the response scheme to form what is often called a “no-choice task”. In a two-alternative, no-choice task, the participant has several levels of confidence with which to respond. For example, in a choice of whether an item has been seen previously, a person may be able to respond “probably new”, “maybe new”, “maybe old”, and “probably old”. Although choices are not directly expressed, the first two responses imply the choice of “new”, and the last two responses imply the choice of “old”. Fitting with SDT, memory models often carry the assumption that to provide these responses, people divide the continuum of memory strength into four parts using three criteria.

Confidence is a useful measure of cognitive activity, but it remains poorly understood. After decades of research, its connection to decision-making processes is still
debated. SDT assumes that confidence ratings are simply a result of choices with an expanded response set (Macmillan, 2002), but Baransi and Petrusic (1998) argue from experimentation that confidence ratings require an extra step in the decision which only partly overlaps other steps.

Baransi and Petrusic (1998) showed that when human subjects made choices for which speed was stressed, the time between the primary decision and confidence response decreased with increasing confidence. When they made choices for which accuracy was stressed, this time was unrelated to confidence level. For accuracy, a decision is made slowly, so the researchers assert that all confidence judgments (low or high) can be finished before or simultaneously with the primary decision. Conversely, when a decision is made quickly, the primary decision takes priority, and additional processing of confidence must take place afterward.

In a choice and confidence task, Baransi and Petrusic (1998) also demonstrated a negative relation between confidence and task difficulty. People can often tell when a decision is objectively difficult and feel less certain of their final choice. Others have confirmed this relation (e.g. Peirce & Jastrow, 1884; Vickers, 1979), but Petrusic and Baransi (2003) showed that it is not entirely monotonic.

Perhaps the least surprising relation with confidence is the one for choice accuracy. Although there is a tendency for overconfidence in most tasks (e.g., Budescu, Wallsten, & Au, 1997; Klayman et al., 1999; Griffin & Tversky, 1992; Gigerenzer, Hoffrage, & Kleinböltting, 1991), and particular experimental manipulations cause confidence and accuracy to diverge (e.g., Dawes, 1980; Lichtenstein & Fischoff, 1977; Merkle & Van Zandt, 2006), we can expect confidence to be some semblance of the probability of being correct. Many times over, researchers have shown that accuracy
and confidence are positively related, even when the difficulty of the stimuli is controlled (e.g., Baranski & Petrusic, 1998; Dougherty, 2001; Garrett, 1922; Merkle & Van Zandt, 2006).

A third relation between the variables discussed so far is a negative one between confidence and decision time. When people are given ample time to make decisions and encouraged to be accurate, fast decisions are typically observed with high confidence (e.g., Baranski & Petrusic, 1998; Henmon, 1911; Festinger, 1943; Van Zandt & Maldonado-Molina, 2004; Vickers & Packer, 1982). The explanation provided by most theorists is that for cautious decision-makers, decisions with a high quality of evidence will quickly home in on the best alternative. This is not the case when speed is emphasized (Vickers & Packer, 1982) and decision-makers are less cautious.

Below, I summarize the expanded Poisson race model, the RTCON model, and the 2DSD model, which each try to explain some of these relations between stimuli, accuracy, confidence, and response time.

### 1.2.1 Expanded Poisson Race Model

A major advantage of the Poisson model is how it has been adapted to account for confidence (Merkle & Van Zandt, 2006; Van Zandt, 2000; Van Zandt & Maldonado-Molina, 2004). Along with choices and response times, it predicts confidence ratings in simple decision tasks. To do this, researchers can incorporate Vicker’s (1979) balance-of-evidence hypothesis from his accumulator model for which confidence is the difference between the evidence levels at the time a decision is reached. His model has not been fully evaluated for explaining confidence and response time, but the expanded Poisson model has. For the Poisson model, this means that $C =$
\[ X_A(t_D) - X_B(t_D) \], where \( C \) is confidence, \( X \) is evidence, and \( t_D \) is decision time (Van Zandt, 2000). The probability distribution of \( C \) can be found using this function of evidence.

The expanded Poisson model predicts that confidence and accuracy are positively related because the correct and incorrect counters both accrue more evidence before incorrect choices, leading to lower confidence. The model also predicts that confidence decreases with task difficulty because stimuli that are difficult to discriminate have smaller differences in their accrual rates, resulting in smaller differences in evidence. The model further predicts that when accuracy is emphasized, people will show a negative relationship between confidence and decision time. This is because longer decisions allow more opportunity for a non-winning counter to accrue evidence.

### 1.2.2 RTCON Model

One disadvantage of the diffusion model that is hard to overcome is its relative stopping rule. In that model, a decision is always made when evidence for one alternative exceeds the others by an amount equal to the difference between the process’s starting point and threshold. Evidence for one alternative is perfectly negatively correlated with evidence for the others at the time a choice is made.

Vickers (1979) pointed out that the sequential probability ratio test model developed by Wald (1947), which is a type of random-walk model, can transform a posterior “belief” about a choice into a subjective probability, but that this subjective probability will be the same for every decision because the thresholds determine the amount of evidence that will favor an alternative at the time of every choice. Thresholds are assumed to not vary while the context of the decisions remains the
same, so confidence will also not vary. If we apply the balance-of-evidence hypothesis to a diffusion model, the resulting confidence will not change between trials of an experiment.

Ratcliff and Starns (2009) propose the RTCON model without a balance-of-evidence for confidence, suggesting that information is not accessible as it accumulates. They draw heavily from SDT to explain confidence and response time. Their focus is on recognition memory, in which the degree of match between a single item and memory for an item is represented as a distribution of stimulation. Criteria divide the distribution into confidence areas. Although the primary decision in the research is between two alternatives, whether an item is old or new, the theory treats each confidence rating as its own alternative. The sizes of the confidence areas determine the drift rates for many separate diffusion processes.

As to the negative relationship between confidence and response time, like other researchers, Ratcliff and Starns (2009) rule out the suggestion that confidence is a direct function of response time (Volkmann, 1934; Audley, 1960). Their model suggests that when an item’s distribution is shifted to the low or high side of the memory strength continuum, a larger portion of the distribution falls into a high-confidence region near the ends of the scale, “certain new” or “certain old”. This creates a larger drift rate for a high confidence rating. Because the focus of the RTCON model is on explaining in what is called a “receiver operating characteristic” of the relations for particular types of responding, the way it accounts for other relations between accuracy, task difficulty, confidence, and response time is unclear.
1.2.3 2DSD Model

Pleskac and Busemeyer (2010) proposed the two-stage dynamic signal detection (2DSD) model to explain separate choice and confidence. Like the RTCON model, it is based on a diffusion model, which forms the first stage for the choice. To model confidence, the authors assume that evidence continues to accumulate in the second stage after a choice. After a fixed time $\tau$ following the primary decision, evidence from the choice process is categorized into confidence ratings. The additional time provided by $\tau$ allows the process to wander away from the decision threshold, where the space is divided by confidence criteria. More evidence in favor of one alternative places the process farther in the direction of that alternative, where it will end up in a higher confidence region.

Pleskac and Busemeyer (2010) claim that the 2DSD model explains many relations between confidence, accuracy, stimulus difficulty, and response time. The simple explanation for why confidence and response time are negatively related is provided by drift rates. As drift rates increase, decision times decrease, but evidence levels increase differentially. This also explains the negative relation between confidence and difficulty because difficult-to-discriminate stimuli have smaller drift rates; hence, lower confidence responses are mapped to the lower evidence levels.

A drawback of this model is that it does not address the no-choice confidence task. Instead, it is applied to tasks with separate choice and confidence responding yet makes no attempt to account for the distribution of times between the primary decision and confidence rating. The “inter-judgment” time, $\tau$, is a single parameter that forms an interrogation-type stopping rule for confidence, meaning that something exogenous determines when evidence stops accumulating. An assumption regarding
an endogenous standard or secondary threshold as in the expanded Poisson model could be added to the model. The basic idea of confidence processing after a choice is the same for both models: further collection of evidence from the same distribution that led to an earlier choice allows the model to account for a range of empirical effects of confidence.

1.2.4 Modified Linear Ballistic Accumulator Model

I’ve chosen to focus on the problem of more than two alternatives in a decision task. The models presented earlier, the accumulator model, the Poisson race model, the diffusion model, and the LBA, all can be used to model decisions regarding more than two alternatives, yet it is easier with the Poisson model and the LBA and notoriously difficult with the diffusion model. Among the researchers who’ve modeled such tasks (e.g., Busemeyer & Townsend, 1992), Usher and McClelland (2004) were motivated to create their leaky, competing accumulator model to explain particular context effects in the decision-making literature. One of the benefits of the LBA is that it was developed from the leaky competing accumulator, so its ability to model multiple choices has already been demonstrated (Brown & Heathcote, 2008).

Although the LBA has successfully handled choices with multiple alternatives, it has not yet been augmented to account for confidence in choices. Because it is the simplest complete model of choice and response time which still does not address confidence or movement planning for responses, I will modify it to do so. In the chapter that follows, I describe this extension of the LBA.
1.3 Non-Decision Times

As stated earlier, the overarching goal of this research is to account for multiple components of a decision task. So far, I’ve covered mathematical models for choice probabilities, response times, and confidence. My proposed model should handle all three variables, but there is a fourth: movement time.

Often, experimenters count the time between stimulus onset and the press of a button for a choice as response time, but they recognize that many processes may be operating during that time. A participant first has to notice the stimulus and attend to it, then perhaps decode it, then make a decision about it. When the decision process is finished, the participant has to map the choice to a response, then program the motion for the response, then execute the motion for the response. Typically these activities take far less time than the decision and are not considered carefully in decision studies. I have even run studies of this character, when the various processes are undifferentiated by experimental methods.

Perhaps a bigger problem with the non-decision components of decision tasks is how they are modeled. In dynamic choice models, these components are typically rolled together into a single term expressed mathematically as a parameter. For example, in a footnote Brown and Heathcote (2008, p. 154) wrote, “We assume that [response time] is the sum of decision time plus a constant extra time, representing all the processing that does not involve decision making, such as time for perception and response production.” They symbolize the non-decision processing time as \( t_0 \), a free parameter added to the decision times of their LBA.

Ratcliff has given the non-decision components more consideration in his diffusion model (e.g., Ratcliff & Smith, 2004). The non-decision parameter \( T_{cr} \) had been a
minor piece of the model until he made it random across trials of an experiment. Drawing it from a uniform distribution with a range from 0 to $s_t$ can improve model fit. However, $T_{er}$ has no particular psychological meaning, though I assume it is the output of a cluster of processes.

The dynamic choice models I’ve reviewed assume that non-decision processes are independent of the decision process and simply add a non-decision component to decision time for fitting response times. This may not always be reasonable. Imagine a task in which people categorize images. If one category contains more complicated images, a person may take longer to attend to all aspects of the images and also longer to make decisions about them.

As a more pertinent example of the dependency between movement time (MT) and decision time (DT), consider a no-choice/confidence task with six confidence ratings arranged as a row of identical keyboard buttons. Decision processing and movement might not be empirically separable in this settings. A sensible response strategy for human subjects would be to hover over the middle two buttons while making a decision and move a finger to the appropriate button after making a decision. Confidence ratings at the ends of the row will take longer to reach than confidence ratings near the middle of the row, but the confidence process will take longer for the ratings near the middle of the row than ratings at the ends of the row. In this case, MT and DT will both be related to confidence, although it has been assumed that MT and DT are independent (e.g., Donders, 1868; Quick, Duncan, & Malcolm, 1962). For easy choices, DT should decrease, but MT should increase. (Another response strategy is to have a separate finger hover over each of 10 or fewer buttons until a
decision is made, but even then MT and DT will be related, though for a different reason).

By having DT and MT model parameters that can compensate for each other, there may be artificial variation in one parameter when restraining another. Because actual MT and DT are both part of response time, if MT is considered a constant within the non-decision time of a choice response time model, parameters for DT in the fitted model will have to make up the difference when actual MT varies. There will be a clear pattern in MT and confidence when farther buttons in a row are for higher confidence responses. Higher confidence is often associated with easier, faster decisions, but in this case it is also associated with longer movements. So long as the MT parameter cannot vary systematically with confidence, parameters of the decision process will increase predicted DT so as to fit the observations of response time.

One way to work around this problem is to remove the movement from the response time period by redesigning the experiment. However, movement may still have an influence on reaction time (RT). Because I choose to focus on MT in this project, I will explain more below. A person asked to select an action from a set when the task is of a probabilistic nature is faced with a decision under uncertainty. The person perceives a probabilistic structure which impacts both decision making and movement.

1.3.1 Movement Times

Actions that require movement of a body part toward a target are often called “aimed” movements. In many laboratory tasks, people make responses on computer hardware, so typically they aim a finger toward a button. Both temporal and spatial
accuracy of aimed movements are important, but some tasks emphasize timing of movement, while others emphasize location of movement.

Woodworth (1899) was the first to study aimed movements. Because he conducted his experiments long before modern electronics became available, the movement tasks he used required repetitive line-drawing to a target, while speed was varied by the pace of a metronome. Tasks were performed over various distances, with right and left hands, and with eyes open and closed to discover how speed and accuracy are affected by these factors. Unsurprising to modern psychologists was the finding that as speed increased, accuracy decreased. This and other results of his work have stood up to scrutiny for over a hundred years.

Thus we see that the speed-accuracy trade-off does not just occur in decision making. As Woodworth (1899) discovered, simple movements display this type of trade-off, whereby faster movements are less spatially accurate. Errors of movement most often occur when people try to perform aiming tasks too quickly. Because high accuracy becomes increasingly difficult to attain as the target of one’s movement becomes smaller or farther away, the speed of movement has a connection to target size and distance.

**Fitts’ Law for Rapid Aimed Movements**

In 1954, the speed-accuracy relation for human motion was expressed as a formal mathematical law by Paul Fitts. The experiments by Fitts involved a repetitive tapping task in which participants moved a stylus back and forth between two target plates as quickly as possible. He theorized that a participant’s movement involved the transmission of information through the nervous system. He drew on information theory (Shannon & Weaver, 1949), in which “information” is a measure for uncertainty,
to calculate the information capacity of the motor system. He defined information capacity as the system’s “ability to produce consistently one class of movement from among several alternative movement classes” (Fitts, 1954, p. 381). With more alternatives, executing a particular movement involves more capacity and is therefore more difficult. His famous equation for the “index of difficulty” of a movement, \( I_d \), uses a base-two logarithm. For a ballistic movement, it is customarily written as

\[
I_d = \log_2 \left( \frac{2D}{W} \right),
\]

(1.1)

where \( D \) is the distance from a starting point to the center of a target and \( W \) is the width of a target as measured in the same direction as the distance.

When Fitts increased the distance between the targets, MT increased. When he increased the width of the targets, MT decreased. MT was directly related to the log-based index of difficulty by the linear equation: \( MT = \alpha + \beta \times I_d \). The terms \( \alpha \) and \( \beta \) are positive empirical constants, parameters needed for the line to fit the observed data. Fitts’ Law is given as

\[
MT = \alpha + \beta \times \log_2 \left( \frac{2D}{W} \right).
\]

(1.2)

The equation tells us that width and distance can each be multiplied by a constant, and MT will not change. In the original study, average MT was nearly identical for targets 16 inches apart with widths of 1/2 inch and targets 8 inches apart with widths of 1/4 inch.

The value of \( \alpha \) may be interpreted as the MT when \( I \) is zero (Welford, 1968), which occurs in a ballistic movement task when width and distance are equal. Typically, \( I \) is between two and seven in laboratory experiments using ballistic tasks. The value of \( \beta \) is the slope of the line. It describes how MT changes as a function of difficulty.
Experimentally, $\beta$ varies based on which motor effector is used because different body parts show different sensitivities to the index of difficulty.

Although researchers have questioned whether Fitts’ Law applies to most movements (e.g., Schmidt & Lee, 2005), Fitts and Peterson (1964) showed that it does apply to quick, discrete movements aimed at a target. When people are asked to make a single movement as quickly and accurately as possible, the equation fits the data well. Jagacinski and Monk (1985) even showed that the law holds for moving a cursor on screen with a head-pointing device.

**Linear Speed-Accuracy Trade-Off for Aimed Movements with Time Constraint**

Fitts’ Law provides a single general principle of motor performance; however, several modifications to the index of difficulty have been made to achieve slightly better fit for particular circumstances (See Plamondon & Alimi, 1997). Where Fitts’ Law has faced the toughest challenge is in movements with constrained time. While many laboratory tasks involve people moving very quickly toward targets, others involve moving at a specific pace toward targets. Participants may be given a goal MT so that accuracy of timing and distance are equally emphasized. Participants receive feedback when responses are too fast or too slow. This type of task displays a different speed-accuracy trade-off. MT is treated as an independent variable, while width is treated as a dependent variable. Variability of movement distance, “effective target width”, has a linear relationship with mean movement distance when all movements are expected to take the same amount of time. In other words, spatial accuracy decreases linearly as speed increases in a constrained MT task (Schmidt, Zelaznik, Hawkins, Frank, & Quinn, 1979). Schmidt and Lee (2005) refer to effective target
width as $W_e$, the standard deviation of movement distance, and express its relation to the other variables as

$$W_e = \alpha + \beta \left( \frac{2D}{MT_G} \right), \quad (1.3)$$

where $MT_G$ is the goal movement time. If the MT goal is changed, but movement distance is maintained, a linear speed-accuracy trade-off still appears (Abrams, Meyer, & Kornblum, 1989).

The linear trade-off contradicts the trade-off expected by Fitts’ Law, albeit for a special kind of movement. One hypothesis is that the logarithmic trade-off will occur when movement follows a feedback-corrected plan, and the linear trade-off will occur when movement follows a premade program. Movement with corrections from sensory-motor feedback is often characterized as “closed-loop”, and movement without corrections is often characterized as “open-loop” (e.g., Schmidt & Lee, 2005). If a person is told to make a movement as quickly as seems reasonable, he will typically follow a closed-loop plan. If a person has only a fraction of a second to make a movement, he will typically follow an open-loop plan because he doesn’t have freedom to make corrections.

Experimental tasks for aimed movements with goal MT usually constrain MT to a very small duration (e.g., 200 milliseconds in Schmidt et al., 1979). However, a linear trade-off may simply be characteristic of the MT-goal paradigm (Wright & Meyer, 1983). This hypothesis was supported by an experiment by Carlton (1994), who observed the number of submovements within longer aimed movements. In one condition, participants moved as quickly as possible toward targets, and in the other condition, participants moved toward targets at a pace set by MT goals. The target width in the former condition was closely matched to the effective target width in
the latter condition. The movement distances were also adjusted so that MTs closely matched between conditions (approximately 400 milliseconds). The MTs were long enough for people to gather visual feedback and correct their movements, but such corrections were fewer in the latter condition. In the simple aiming condition, nearly all movements consisted of submovements, but in the goal-MT condition, less than 20% of movements consisted of submovements.

The nervous system has noise in both afferent and efferent nerves. Just as SDT recognizes variability is inherent to sensation of stimuli, variability is inherent to motor responding. To accommodate the noise in the generation of movement, Crossman and Goodeve (1983) proposed the “iterative-corrections” model of human motion. In 1963, they suggested that Fitts’ Law could be derived mathematically from assumptions about feedback-controlled movement. They assumed that aimed movement has two phases: a ballistic phase and a correction phase that together home in on a target. The two phases alternate quickly in each submovement. The first phase operates for a fixed duration and has some inaccuracy associated with it. A feedback process in the second phase detects the error and adjusts the size and direction of the next submovement. Under this theory, MT is determined by the number of iterations needed to reach the target.

To generate reasonable MTs, the iterative-corrections model required many corrective iterations, but some researchers (e.g. Langolf, Chaffin, & Foulke, 1976; Wright & Meyer, 1983; Meyer, Abrams, Kornblum, Wright, & Smith, 1988) analyzed rapid aimed movements and concluded that participants made fewer corrections within them than were predicted by the model. Another problem for the model came from assuming constant time for each phase (e.g., Schmidt et al., 1979; Meyer et al., 1988),
but the main criticism was for its theoretical deviation from Woodworth in assuming that all submovements are structurally the same. Woodworth (1899) had argued that the initial impulse of movement was qualitatively different from the remainder of movement.

Schmidt et al. (1979) developed a theory of impulse variability to explain the linear speed-accuracy trade-off in time-constrained movements. However, other researchers (e.g., Jagacinski, Repperger, Moran, Ward, & Glass, 1980; Meyer et al., 1990) attributed the logarithmic trade-off in other aimed movements to the initial impulse of movement and later corrections. The generally accepted theory for the two types of aimed movement is the optimized-submovement model (Meyer et al., 1988), which combines some ideas from both paradigms.

**Optimized-Submovement Model for Rapid Aimed Movements**

Meyer et al. (1988) explained numerous results in the motor performance literature by developing a model with basic assumptions like those of the iterative-corrections model. In their theory, aimed movements consist of a ballistic phase and a feedback-control phase. The phases cycle between submovement and correction until the effector reaches the target. As in the iterative-corrections model, all submovements have the same structure so that the first phase is structurally preprogrammed but parameters are programmed at the outset of movement. Its description is based on the impulse-variability model so that a linear speed-accuracy trade-off occurs within the first phase. When the ballistic submovement of the first phase is inaccurate, the second phase corrects by setting new parameters for the next submovement to be programmed. This second phase requires active processing of visual feedback about movements.
The optimized-submovement model (OSM) has a closed-loop process operating as an iterative open-loop process. It explains the logarithmic speed-accuracy trade-off for movements with non-constrained time by the series of corrections made to submovements. A person attempts to optimize MT through adjustments to the initial impulse of movement and later submovements. Meyer et al. (1988) showed that two submovements reduce the average total MT while maintaining spatial accuracy compared to a single movement. They supposed that typically only a single secondary submovement is needed when the primary submovement overshoots or undershoots the target and therefore developed a square-root approximation of Fitts’ Law:

\[
MT = \alpha + \beta \times \sqrt{\frac{D}{W}}.
\]

The dual-submovement model fits experimental data well, but the authors noted that some combinations of movement distance and target width generated more than two submovements. The optimized-submovement model suggests that each additional submovement decreases MT, but the marginal decreases are diminishing rather than constant. Meyer et al. (1990) reexamined the data with a different kinematic analysis and decided that three submovements also frequently occurred. They revised the model to allow for many submovements and found that the best model fits were achieved for three and four submovements. Their model can be written as

\[
MT = \alpha + \beta \left( \frac{D}{W} \right)^{1/n},
\]

where \( n \) is the maximum number of submovements. As this parameter increases, the model approaches Fitts’ Law. Fitts’ Law is a special case of the OSM. Furthermore, when \( n = \infty \), the equation becomes \( MT = \alpha + \beta \log_e \left( \frac{D}{W} \right) \). For a single submovement,
the relation between $MT$ and $D/W$ is linear because the equation becomes $MT = \alpha + \beta \left( \frac{D}{W} \right)$ when $n = 1$.

Although the OSM is popular, there is always some caution to endorsing a model. In this case, it lies with the prediction that submovements show a linear trade-off between average velocity and endpoint variability. If corrections drive a logarithmic trade-off in aimed movements and any submovements are so short as to not have any corrections within them, submovements will show a linear trade-off instead. Although movements with MT goals show a linear trade-off (Schmidt et al., 1979), there is room to debate whether single submovements in larger movements without MT goals show it. In the 1990 revision, the authors cite the experimental results of Searle and Taylor (1948) to support their claim for the OSM that the initial phase of aimed movements have a linear trade-off, but qualitatively, this is not true. Similarly, Liao, Jagacinski, and Greenberg (1997) showed a negatively accelerated curve relating spatial variability and average velocity of submovements and developed a stochastic model incorporating that function.

1.3.2 Movement Programming Times

The literature on MT shows that it is not constant for movements to different targets at different distances, so choice response time data may be “contaminated” by MT. In an experimental paradigm with multiple response options, the response targets are often different distances from the participant. This may occur for decisions with more than two choice alternatives or decisions with multiple confidence ratings. To avoid this confound, MT may be removed from response time through experimental
methods. Instead of measuring the time from stimulus onset until the press of a button, researchers might measure the time until movement begins toward a button.

Unfortunately, if MT is extracted from response time, there may still be contamination of RT. RT reflects non-decision processes other than MT, such as perception. Specifically, RT might still contain movement programming time (MPT). The models of choice response time that I’ve reviewed don’t account for MT appropriately, let alone MPT. This is a concern because MPTs may be significant. If for any reason they vary systematically counter to decision processing times, modeling the decision task becomes more difficult.

The implication of the OSM is that behavior under Fitts’ Law is the product of repeated movement corrections. If the corrections are viewed as reprogramming of the original movement program (Laszlo & Livesey, 1977), the initial programming of movement may also conform to the OSM. The literature has several studies which support a simple relation between MT and MPT and several that suggest no relation exists between them. Because I will invoke the OSM for explaining MPTs in Chapter 2, this topic is of particular interest.

Support for the idea that MPTs should vary in a manner similar to the MTs that follow them comes from Fitts and Peterson (1964). They found that in choice RT, the distance and width of targets had an effect on RT along with an effect on MT. There was a statistically significant relationship between RT and index of difficulty. However, when regressing RT on index of difficulty, the slope was much smaller than the slope for MT in the same experiment.

Duarte and Latash (2007) examined the effect of distance and width on the duration of posture adjustments by people performing foot-pointing tasks while standing.
The effect on posture adjustment time prior to movement of their feet was very similar to that on MT of their feet. The conclusion was that Fitts’ Law applied to movement programming before and during movement.

Rosenbaum, Halloran, and Cohen (2006) showed that movements are programmed/planned in great detail. The researchers wanted to know whether people prepare movement starts in more detail than movement ends. By the iterative-corrections model and OSM, one could imagine that the control phase with its corrections to movement allows the movement phase before it to be less carefully prepared. This would mean aiming of movements becomes more precise at each step, but the experimental results contradict this notion. Participants performed a grasping movement to lift a plunger from one ring to another ring. Assuming that removing and inserting the plunger into equally sized rings requires equal accuracy, varying the size of a ring could produce different effects depending on whether the variable ring is at the start or end of movement. If movement preparation focuses more on the start of movement, the ring from which the plunger is lifted should have more influence on the grasps people use. However, the researchers found that widths of the start and end rings equally affected participants’ grasps. The rings could both be considered targets of movement with equal weight in determining the characteristics of movement.

Although the experiment by Rosenbaum et al. (2006) did not measure MPT, it seems reasonable that MPT, like grasp, would be affected by the entire movement to be programmed. MPT and grasp both involve preparation for motion, so they both might change when the goals of movement change. One line of reasoning for how MPTs could be proportional to MTs involves the possibility that the number of submovements is not very sensitive to the distance of movement. The same number of
submovements might be used over a range of distances and still be optimal, so that the submovements are longer when their overarching movements are longer. If the initial submovement lengthens with the total movement distance, the programming time for the first submovement increases with MT because the difficulty of each increases.

Indeed, other researchers have shown that factors affecting MT also affect MPT. MT and MPT change very similarly with changes in movement distance and somewhat similarly with changes in target size (Juras, Slomka, & Latash, 2009). However, Klapp (1975) stated that although RT might often be impacted by target distance, it is not impacted by target width at long distances (> 335 mm). Because RT is impacted by target width at short distances when the MTs are small, he hypothesized that very short movements are fully programmed during RT, but longer movements are controlled by feedback-based, ongoing programming. In one experiment, elimination of visual feedback had a greater decrement to accuracy for longer movements. (This research indirectly influenced the theory behind the OSM.)

 Some studies do not show any changes in MPT when MT changes. One study comes from Drury and Corlett (1975), who had people discriminate stimuli. The participants were to touch buttons in response to lights arranged on an arc of a circle. When a light turned on, the participant touched the button below it. The index of difficulty was calculated differently from usual because participants did not return to a starting point after each trial, so their study is less applicable to the choice RT paradigm. Nevertheless, the researchers found almost no relation between RT and index of difficulty.

While studying two-dimensional movements as they relate to Fitts’ Law, Jagacinski and Monk (1985) found that the slope of the function relating choice RT to the index
of difficulty was zero. Although $\alpha$ changed under different conditions, $\beta$ did not. Similarly, Liao et al. (1997) showed that simple RTs were unrelated to index of difficulty. They were studying target acquisition by younger and older adults when they found that although the mean RT was greater for older adults, the slopes of the Fitts’ Law function for both younger and older adults were close to zero.

The work of Liao et al. (1997) offers great insights into aimed movement but no definitive answer for whether MPTs are governed by the same principle as MTs. Although slopes for RT functions of index of difficulty were not significantly different from zero, analysis of first submovements showed a logarithmic trend. Later submovements were shorter and possessed more unexplained variability, but variability of first submovements was consistent with Fitts’ Law.

I make the argument that some aspects of a movement even up to the end may be programmed at the start of movement, but programming is mostly for the first submovement. Because there is potential for a logarithmic speed-accuracy trade-off in the first submovement of aimed movement, the mathematical models of MT for aimed movement should also be applied to MPT for aimed movement. Whether MPT within RT is affected by the same variables as MT is an open topic.

Klapp and Erwin (1976) suggested that choice RT might be a logarithmic function of response duration. They found that MT and choice RT have a positive relation but an increase in MT may not be a sufficient or necessary condition for an increase in choice RT.

Fitts and Peterson (1964) had difficulty concluding whether RT and MT were independent of each other, but a recent review of research literature (Glover, 2004) suggests that planning and control of actions are separate processes. This is not
consistent with the OSM, for which we could assume that the movement phase of the first cycle is programmed by the control phase at a point in time before experimenters begin measuring MT and that the same type of control programs later movement phases. Brain imaging and neuropsychology studies indicate that planning the start of a movement and controlling it after it starts use different visual centers of the brain, although they appear to be influenced by similar factors.

Thomas (1971) concluded that people typically cannot control the timing of their movements without affecting the timing of their decisions and vice versa. He gave participants in his experiment goals for MT and DT, and observed them struggle to achieve each one alone. They had even more difficulty achieving them simultaneously when the goals were not the same. Only after extensive practice were the distributions of MT and DT independent.

1.4 Goals of the Research

The research presented here has several goals. The highest goal is to show the extent of influence of the motor component of a task on response time by developing a decision model that accounts for MT and MPT. Dynamic choice models have neglected the variability in non-decision times by assuming a constant baseline for response times with parameters such as $T_{cr}$. However, the full dynamic choice model I’ve created includes expressions for MT and MPT. It predicts the speed-accuracy trade-off in decision making and the speed-accuracy trade-off in aimed movement.

Figure 3.12 shows the trends expected for confidence, MT, RT, and response time in the four conditions of the experiment described in Chapter 3. The relations may not be completely linear or monotonic, but the picture provides the approximate
patterns of predictions. The figure was generated by choosing reasonable parameters of the model described in Chapter 2 and plotting the predicted values of MT and RT for the lowest and highest confidence ratings or nearest and farthest targets. One can see that confidence ratings take longer than choices but that higher confidence ratings take less time than lower confidence ratings. The more important thing to see is that triangular conditions remove the effect of distance on MT and MPT. If an experimenter did not measure MT separately, the effect of target distance on total response time would be large. In the confidence condition, the same effect would run counter to the effect of increasing confidence on RT.

Some related models are formed from the full model I specify in the next chapter. The basic model accounts for non-decision times in the manner of so many other choice response time models — adding a constant. The expanded model accounts for MT and/or MPT with the OSM. I expected first that the basic model should fit RT better than response time because RT does not contain MT, a significant variable for which a typical choice response time model does not explicitly account. Second, when modeling response time, I expected that the basic model should fit better when response times have not been differentially impacted by MTs. Third, when modeling RT (by itself or with MT), I expected that the expanded model should fit better by explicitly accounting for MPT.

The full model should account for several effects in decision making and confidence judgment previously described. The LBA can already predict the shapes of choice RT distributions and effects such as slow errors (Brown & Heathcote, 2008). My model is partly based on the LBA, so it should perform similarly, but it should also account for confidence. A secondary goal of the research is to use confidence ratings to gain insight
into the human decision process. Not only do I collect subjective probabilities, I observe the time course of their materialization. My model predicts a positive relation between confidence and accuracy as well as a negative relation between confidence and RT. RTs should be greater for lower confidence ratings because these flow from more difficult decisions.

In studying decision timing, the operational definitions of DT are often fuzzy. A simple first step to begin measuring what I most want to measure, DT, is to extract MT from response time. As this introduction illustrates, the work combines ideas from decision research and motor performance research. Many multi-choice tasks require that participants indicate their choices by moving to one of several response keys. My work will contribute to the improvement of experimental methods for the n-choice task. Developing the method for the experiment described in Chapter 3 was a challenge, but future implementation (even by other researchers) should be less difficult.

The experiment is meant to isolate the influence of movement difficulty on response times. Related to the other goals is my expectation that I can remove any deleterious effects of MT on response time. Through the use of Fitts’ Law, it should be possible to make all MTs nearly equal for a multiple-alternative, no-choice/confidence task. Theoretically, this also makes MPTs nearly constant. These results can then be compared to others when I do not compensate for Fitts’ Law. Model fits can also be compared for conditions that use Fitts’ Law versus those that don’t.
1.5 Organization of this Dissertation

In the second chapter, I explain the full model to account for choice, confidence, RT, MT, and MPT. Most of the mathematical description serves to clarify the motivation for model details.

In the third chapter, I describe the experimental task to be modeled. The experimental methods are designed to obtain data useful for evaluating the models, including choice frequencies, confidence ratings, response times, and movement times. The same chapter’s experimental results section includes summaries of the data and checks for expected patterns.

The fourth chapter is devoted to model evaluation. I first describe the fitting method, then provide measures of fit for a set of models related to the full model. I show the quality of fits graphically as well and provide discussion of the models’ comparative performance.

In the fifth chapter, I review the goals of the research and provide general conclusions. I also consider the limitations of this work as well as future directions.
Chapter 2: MODEL DEVELOPMENT

I’ve chosen to modify the LBA for modeling the data from my proposed experiment because I feel that it balances simplicity with sufficiency of explanation. Although it has not been used to model confidence, I adapt it to that purpose.

2.1 Linear Ballistic Accumulator

The LBA has convenient closed-form expressions for response time probability distributions. The probability density function (PDF) of a normal distribution with mean $\mu$ and standard deviation $\sigma$ is written as $\phi(\cdot|\mu, \sigma)$, and its cumulative density function (CDF) is written as $\Phi(\cdot|\mu, \sigma)$. For the standard normal distribution, the PDF is written as $\phi(\cdot)$, and the CDF is written as $\Phi(\cdot)$. Traditionally, lowercase letters refer to PDFs, and uppercase letters refer to CDFs. For a single accumulator $i$, Brown and Heathcote (2008) provide the CDF for the time required to reach the threshold, $b$:

$$F_i(t) = 1 + \frac{b - A - tv_i}{A} \Phi \left( \frac{b - A - tv_i}{ts} \right) - \frac{b - tv_i}{A} \Phi \left( \frac{b - tv_i}{ts} \right)$$

$$+ \frac{ts}{A} \phi \left( \frac{b - A - tv_i}{ts} \right) - \frac{ts}{A} \phi \left( \frac{b - tv_i}{ts} \right)$$

because the starting evidence, $a$, is randomly drawn from a uniform distribution with lower limit 0 and upper limit $A$ and the drift rate, $d_i$, is randomly drawn from a
normal distribution with mean $v_i$ and standard deviation $s$. Naturally, $b$ must be greater than $A$, and $t$ must be positive. If $t < 0$, $f_i(t) = 0$.

The PDF derived from the CDF for finishing time on accumulator $i$ is

$$f_i(t) = \frac{1}{A} \left( -v_i \Phi \left( \frac{b - A - tv_i}{ts} \right) + s\phi \left( \frac{b - A - tv_i}{ts} \right) \right) + \frac{1}{A} \left( v_i \Phi \left( \frac{b - tv_i}{ts} \right) - s\phi \left( \frac{b - tv_i}{ts} \right) \right).$$

(2.1)

Figure 2.1 illustrates the process with a modification that does not change the picture. This equation is the PDF for just one isolated accumulator. Decision times are determined by the minimum finishing time for all accumulators. The probability that an accumulator will finish before competing accumulators is described by a different distribution. The joint probability that accumulator $i$ finishes at time $t$ and that it finishes first is

$$f_i(t) \prod_{j \neq i} (1 - F_j(t)).$$

This joint distribution does not integrate to one. In this case, the total area under the curve instead reflects the probability of the response $i$ when other responses $j$ are available. Another way to compute the probability of response $i$ is to evaluate the function’s associated CDF as $t \to \infty$. The CDF can be found by numerical integration (For a full explanation, see Brown & Heathcote, 2008).

Evidence $z_i$ at time $t$ is simply $a + t \times d_i$. Of course, $a$ and $d_i$ are variable, so the authors also provide the CDF and PDF for the evidence at a given time. The probability of evidence $z$ on accumulator $i$ at time $t$ is

$$w_i(z) = \frac{1}{A} \left( \Phi \left( \frac{z}{t} \middle| v_i, s \right) - \Phi \left( \frac{z - A}{t} \middle| v_i, s \right) \right).$$

It expresses the idea that the probability that an accumulator has evidence $z$ is the probability that its rate will cause it to accumulate a new amount of evidence from
the minimum possible distance between a starting point $a$ and $z$ to the maximum possible distance between $a$ and $z$ times the probability of each appropriate value of the starting point. It was derived from the following equation:

$$W_i(z) = \frac{ts^2}{A}\left(\phi\left(\frac{z}{t}\bigg|v_i,s\right) - \phi\left(\frac{z - A}{t}\bigg|v_i,s\right)\right)$$

$$+ \frac{1}{A}\left((A - z + v_it)\Phi\left(\frac{z - A}{t}\bigg|v_i,s\right) + (z - v_it)\Phi\left(\frac{z}{t}\bigg|v_i,s\right)\right).$$

Again, the probability for one accumulator is in isolation from other accumulators.

### 2.2 Ballistic Accumulator for Confidence

I made several modifications to the LBA for modeling data from the experiment described in Chapter 3. I allow the decision threshold to vary between conditions of the experiment, I constrain accumulation rates to sum to a constant, and I determine the accumulation rates from independent variables of the experiment. Most importantly, I define confidence based on the evidence in the accumulators.

#### 2.2.1 Decision Thresholds

Differences in observed response times between forced-choice tasks and no-choice confidence tasks are typically attributed to changes in the decision threshold. The model will produce confidence response times that are longer than choice response times when more evidence must accumulate for confidence response times, that is, when the threshold for no-choice confidence is higher than for forced choice. I assume that for a combined forced-choice/confidence task, two thresholds would be required: a lower threshold for the initial decision and a higher threshold for the confidence judgment.
Petrusic and Baranski (2003) showed in a forced-choice/confidence task that response times increase when confidence is measured. The Ballistic Accumulator for Confidence (BAC) should fit this pattern of data by increasing the decision threshold for forced-choice/confidence tasks over the decision threshold for forced-choice/no-confidence tasks.

With two thresholds, there can be processing time between evidence reaching one threshold and then the other. Consistent with this idea, Van Zandt and Maldonado-Molina (2004) showed experimentally that people reverse their responses without any change in the context of their decisions. In a recognition memory paradigm with a forced-choice/confidence task, participants made choices and then gave confidence ratings to show that they had changed their minds on a small portion of trials. The authors argued, “Only a process that continues to unfold over time […] could result in the awareness that a response was incorrect.” Their expanded Poisson model reflects this belief to accommodate their data. I mention this assumption because it may be very important in fitting of my proposed model.

Usher, Olami, and McClelland (2002) proposed that in the context of dynamic choice models, increases to the decision threshold serve another purpose. The Hick-Hyman Law (Hick, 1952; Hyman, 1953), which resembles Fitts’ Law, states that choice RT increases logarithmically with the number of alternatives. The Hick-Hyman Law can be explained by increases to the decision threshold as the number of alternatives increases, and confidence ratings can be seen as a way of increasing the number of alternatives in a choice. This means that when fitting the model, the threshold, as a free parameter, might be higher for no-choice confidence tasks than for choice tasks.
2.2.2 Accumulation Rates

Another way to accommodate the Hick-Hyman Law was discussed by Brown and Heathcote (2005). Forcing the accumulation rates to sum to a constant as if there is a limit on processing capacity will shrink each $d_i$ as the number of accumulators increases, thereby increasing the finishing times. I explain the use of this model feature more in Chapter 4, where it serves to maintain the proper speed of decision-making across contexts rather than to mimic the Hick-Hyman Law.

When my model is applied to the experiment in the next chapter, each accumulation rate $d_i$, relies on the exponential distribution parameter $v_i$, but $v_i$ is not a free parameter. Each $v_i$ is the inverse ratio of the probability that the $i$th choice alternative is right for a stimulus. It is the reciprocal of a simple likelihood ratio which will produce an average $d_i$ equal to the likelihood ratio. The model predictions are based firmly on the known variables in the experiment.

2.2.3 Evidence for Confidence

Van Zandt (2000) used Vicker’s (1979) balance-of-evidence hypothesis to account for confidence. Within the expanded Poisson model, the balance of evidence explains the relations among response time, accuracy, and confidence. When response thresholds are sufficiently high, a negative relationship between confidence and decision time emerges. Decision time will increase when the winning counter has a reduced accumulation rate while a non-winning counter maintains its typical accumulation rate, resulting in a smaller difference between the two counters.

Similar predictions are made by incorporating a relative balance-of-evidence hypothesis (Merkle & Van Zandt, 2006). This adaptation of the Poisson model relates
to support theory (Tversky & Koehler, 1994) and states that a person’s confidence is equal to the evidence for the winning counter divided by the total evidence at the time of a decision. Because the evidence for the winning counter is at its threshold when it wins, the expression can be simplified as $K_{\text{win}}/(K_{\text{win}} + X_{\text{lose}}(t_D))$. This formulation has the advantage of placing confidence on the probability scale that is often used by participants.

Because evidence increases by large discrete units in the Poisson model, inappropriate distributions of confidence may emerge; many values of confidence cannot occur when thresholds are low. To work around this problem, variability in the thresholds between trials is added to the model (Merkle & Van Zandt, 2006) or a Gaussian kernel may be used to develop a continuous-scale density function. Pleskac and Busemeyer (2010) suggested perturbing confidence with noise and partitioning the distribution into ratings in much the same way as signal detection theory. Modifying the LBA for confidence avoids this problem altogether because evidence is on a continuous scale.

With expressions from the LBA giving the probability density for the evidence levels, I can define confidence, $c$, using the relative-balance-of-evidence hypothesis. Momentary confidence for one of multiple accumulators is

$$c_i(t) = \frac{z_i(t)}{\sum_{j=1}^{N} z_j(t)}, \quad (2.2)$$

where $N$ is the number of accumulators. The LBA is not suited to this expression of confidence because it allows evidence to be negative on any accumulator. Negative evidence on one accumulator makes it possible to have confidence above one because the sum of evidence may be less than evidence for accumulator $i$. It also makes it possible to have confidence for a choice below zero when the sum of evidence is negative. In that case, the alternative with the most negative evidence would produce
Figure 2.1: Modified LBA. The diagram shows how the parameters of the BAC produce a decision time, $t_d$, for a single accumulator. The starting point, $a$, is from a uniform distribution. Evidence, $z$, accumulates over time until it reaches threshold $b$.

the highest confidence if chosen. This problem arises because the accumulation rates are drawn from a normal distribution.

To avoid the problem of negative evidence, I keep the starting points and accumulation rates non-negative. I use an exponential distribution of accumulation rates, where the PDF of $d_i$ is $v_i e^{-v_i d}$ when $d_i$ is non-negative and $v_i$ is a non-negative rate for the distribution. Figures 2.1 and 2.2 illustrate this modified version of the LBA, which I call the ballistic accumulator for confidence (BAC).
Figure 2.2: Drift rate distribution. Drift rates for a single accumulator, $d_i$, are drawn from an exponential distribution with rate parameter, $v_i$, to describe its shape. The plot shows an example where $v_i = 1$. 

When $v=1$, $p(d) = e^{-d}$. 

- density 
- accumulation rate $d$
The CDF for \( z \) at time \( t \) with this new assumption for accumulation rates is
\[
S_i(z) = \begin{cases} 
1 + te^{-v_i z/t} \left(1 - e^{v_i A/t}\right) \frac{1}{v_i A} & \text{if } z \geq A \\
(-t \left(1 - e^{-v_i z/t}\right) + v_i z) \frac{1}{v_i A} & \text{if } 0 < z < A \\
0 & \text{if } z < 0.
\end{cases}
\] (2.3)

The PDF for \( z \) at time \( t \) is the CDF differentiated with respect to \( z \), so
\[
s_i(z) = \begin{cases} 
-e^{-v_i z/t} \left(1 - e^{v_i A/t}\right) \frac{1}{A} & \text{if } z \geq A \\
(1 - e^{-v_i z/t}) \frac{1}{A} & \text{if } 0 < z < A \\
0 & \text{if } z < 0.
\end{cases}
\] (2.4)

Somewhat simpler equations for the probability distributions of decision times can be calculated. The CDF for finishing times looks very similar to the first piece of the CDF for evidence \( z \) at time \( t \). It is the complement of the probability that evidence is below \( b \) at time \( t \):
\[
G_i(t) = -te^{-v_i b/t} \left(1 - e^{v_i A/t}\right) \frac{1}{v_i A}.
\] (2.5)

Differentiating with respect to \( t \) gives the PDF for finishing times:
\[
g_i(t) = \left((v_i b + t)(e^{-v_i (b-A)/t} - e^{-v_i b/t}) - v_i A e^{-v_i (b-A)/t}\right) \frac{1}{v_i A t} \] (2.6)

When I replace the normal distribution of \( d_i \) with the exponential distribution of \( d_i \), these functions replace the LBA functions.

Using the above distributions to find the joint probability distribution of choice, confidence, and RT requires numerical integration; however, it can be quickly approximated by simulation. I place the explanation of the above equations in the appendix for the curious reader because they will not be used for fitting the models to the data.

### 2.3 Modeling Non-Decision Processes

Adapting the LBA for confidence created the BAC, but it is only part of the model for this study. I set out to see whether it is reasonable to use a constant in modeling
non-decision times in a decision task. To this end, a model with a simple constant will be compared to a model with a more elaborate account of non-decision times.

In expanding the BAC to include MTs and MPTs, the number of parameters must increase, threatening to make the model’s explanation of data incomprehensible. Adding parameters adds flexibility so that knowing what drives predictions becomes difficult. This happens in the RTCON model (Ratcliff & Starns, 2009), for which 24 free parameters fit the data from a task with six confidence levels (such as five criteria and six drift rates). With more levels of confidence and more choice alternatives, the model would need many more independent diffusion processes with additional free parameters for each decision task. I attempt to limit my model to the parameters in Table 2.1, and set $v$ according to the context of the experiment rather than allow it to freely vary.

### Table 2.1: The parameters of the full model. Several nested and/or related models will be evaluated.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b$</td>
<td>decision threshold</td>
</tr>
<tr>
<td>$A$</td>
<td>range of starting evidence</td>
</tr>
<tr>
<td>$v$</td>
<td>exponential rate of accumulation rate</td>
</tr>
<tr>
<td>$s_m$</td>
<td>standard deviation of MT</td>
</tr>
<tr>
<td>$\alpha_m$</td>
<td>minimum MT</td>
</tr>
<tr>
<td>$\beta_m$</td>
<td>slope of MT</td>
</tr>
<tr>
<td>$\alpha_p$</td>
<td>other non-decision processing time</td>
</tr>
<tr>
<td>$\beta_p$</td>
<td>slope of MPT</td>
</tr>
<tr>
<td>$n$</td>
<td>maximum number of submovements</td>
</tr>
</tbody>
</table>
2.3.1 Movement Times

When Fitts’ Law or the optimized-submovement model (OSM) is fitted to data via linear regression, one may assume that error is normally distributed. Throughout the MT literature, simple least-squares regressions are used for fitting the two models to data, whereby the assumption of a normal distribution is implied. It is implied rather than overtly stated because the actual distribution of MTs is not of interest to many researchers.

If we concern ourselves with the shapes of RT distributions, we should also concern ourselves with the shapes of MT distributions. The equations for these MT models provide the means of normal distributions, so an additional parameter for the standard deviation should be supplied when other fitting methods are used.

Later I will show that Fitts’ Law is inferior to the OSM for explaining the data from the experiment in Chapter 3, so I present only the OSM for MTs within the expanded model. The OSM gives the expected value of MT for a particular combination of distance and width. The full PDF of MT, \( t_m \), is then

\[
\phi \left( t_m \left| \alpha + \beta \times \left( \frac{D}{W} \right)^{1/n}, s_m \right. \right),
\]

where \( s_m \) is a free parameter for the standard deviation of the distribution. \( D \) and \( W \) are specified by the experimenter for the apparatus on which people make confidence responses, so the distribution will be conditional on confidence. The conditional PDF of MT becomes

\[
m(t_m|c) = \phi \left( t_m \left| \alpha_m + \beta_m \times \left( \frac{D_c}{W_c} \right)^{1/n}, s_m \right. \right). \tag{2.7}
\]
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>number of choice alternatives and accumulators</td>
</tr>
<tr>
<td>z</td>
<td>evidence</td>
</tr>
<tr>
<td>c</td>
<td>confidence</td>
</tr>
<tr>
<td>( t_d )</td>
<td>decision time</td>
</tr>
<tr>
<td>( t_r )</td>
<td>reaction time</td>
</tr>
<tr>
<td>( t_m )</td>
<td>movement time</td>
</tr>
<tr>
<td>( t_p )</td>
<td>movement programming time</td>
</tr>
<tr>
<td>( D )</td>
<td>movement distance</td>
</tr>
<tr>
<td>( W )</td>
<td>target width</td>
</tr>
</tbody>
</table>

Table 2.2: Other nomenclature. Some important symbols used in the probability functions are collected.

### 2.3.2 Movement Programming Times

As for MPT, I assume that MPT and DT sum to RT, \( t_r = t_p + t_d \), where \( t_r \) is reaction time, \( t_p \) is movement programming time, and \( t_d \) is decision time (See Table ??). To avoid added complications, I ignore the variability in MPT. The expression for MPT is

\[
 t_p = \alpha_p + \beta_p \times \left( \frac{D_c}{W_c} \right)^{1/n} .
\]  

(2.8)

Of course, there are other non-decision components still unaccounted for. Total response time is the sum of RT, MT, and other non-decision time. Like other decision researchers, I assume that these components of the other non-decision time form a constant extra time. However, I do not include an extra parameter because it would be redundant with \( \alpha_p \), which is inseparable from \( t_r \).

### 2.4 Models to Compare

There are six models to evaluate and five essential comparisons among them. Distinctions among the models can be seen in Table 4.1. The four basic models
add a constant to decision time in computing either total or RT. This may be $\alpha$ to describe all non-decision times in the task or $\alpha_p$ to describe non-decision times which contribute only to reduced RT. The “expanded” set of models has $\alpha_p$ but also incorporates the OSM to account for MTs, MPTs, or both. Because the differences and similarities between models are related to the decision task to be modeled, longer descriptions are in Chapter 4 after the experimental methods have been explained in Chapter 3.
Chapter 3: EXPERIMENT

Many experiments elicit simple choices and confidence ratings from decision makers in a controlled setting. However, they usually measure only the responses and the total time to complete the responses. The experiment I’ve designed breaks the response time down into reaction time (RT) and movement time (MT). Furthermore, where other experimenters have looked for ways that the total response time varies with confidence, I looked for ways that both RT and MT vary with confidence.

The main task participants performed was categorization, sometimes with an added confidence component. I set up the stimuli and response options in a way that is easy to understand but with which participants do not have extensive experience. The task required participants to place numbers into pre-established categories (e.g., Merkle & Van Zandt, 2006). In all cases, participants chose one of four basic responses for categorizing a number, knowing that numbers from each category had a different mean. The categories’ variance made some numbers difficult to correctly categorize. Each number was randomly drawn from one of four univariate, normal distributions which partially overlapped the others.

The model I’ve proposed indicates that RT is related to MT through movement programming time (MPT), but it also indicates that RT is related to confidence. To isolate the influences of MT and confidence on RT, one condition of the experiment
was designed to keep MT relatively stable while confidence varied, and another condition was designed to keep confidence relatively stable while MT varied. Yet another was designed to cause MT and confidence to vary together. To increase and decrease variations in MT and confidence, participants’ method of responding was directed by instructions appropriate to the response apparatus.

Two factors characterize the conditions of the experiment: input display for responses and confidence instructions. The input display was either of a more traditional form for a no-choice confidence task or of a form created with Fitts’ Law in mind. The confidence instructions either required participants to make confidence judgments or they did not.

After explaining the two factors of the experiment’s design in depth, I provide details on the participants, stimuli, apparatus, and procedure of the experiment. The apparatus is the most distinctive feature of the experiment, so it receives the most explanation. After the methods are described, the results are provided. The results section is divided by the four variables that I measured: choice, confidence, reduced response time, and movement time. Finally, there is a short discussion of the conclusions.

3.1 Input Display

One factor in this experimental design is the input display for the computer interface. In trying to account for MT and choice RT, I used what Thomas (1971) called a “combined manual and decision task”. I attempted to create a response motor scheme that feels natural, so I had people point to one response among options on a touchscreen. A touchscreen enables people to interact directly with information
on screen. It does not require any intermediate device that would be held in the hand, such as a mouse, joystick, or stylus. Greenstein and Arnout (1988) highlighted intuitiveness as a primary advantage of touchscreens, but imprecise human interaction with them as a primary disadvantage. However, Sears and Scheiderman (1991) showed that people with varying amounts of computer experience make reasonably accurate and fast finger movements to targets as narrow as 7 millimeters (16 pixels) on a non-stabilized touchscreen. More recently millions of people have begun using touchscreens daily in technology such as smartphones and automatic teller machines.

The choice options in my experiment are arranged so that areas for responses which belong to the same category lie contiguously along a linear path, and higher confidence responses are farther along that path than lower confidence responses. Because spatial accuracy and speed (but not specific timing) of finger movements are emphasized, Fitts’ Law (Fitts, 1954) applies to movements toward these areas. Fitts’ Law states that MT is a function of the ratio of target distance and target width, so for half of the conditions, the widths of the touch areas increase linearly with their distances from the fixation point. In these conditions, response regions for categories have a triangular shape. For the other half of the conditions, touch areas for responses are all of equal width so that response regions for categories have a rectangular shape. Figures 3.3 and 3.4 show rectangular and triangular input displays, respectively.

In Fitts’ original work, distance of movement was on a single dimension, but with my input display for the computer, a person must move a finger in two dimensions. How Fitts’ Law applies to two-dimensional (2-D) movement has been studied by a small number of researchers. Lepine, Glencross, and Requin (1989) studied movement and movement programming in multiple dimensions and concluded that values along
each dimension are independently selected then integrated in a single motor program. MPT for the different dimensions is not additive. When precueing participants in their experiment for the target location on one, two, or more dimensions, they saw nearly no change in timing of movement. If people could program movement for each dimension separately, knowing any coordinate in advance should have hastened their motor responses.

In earlier research on motor programs by Goodman and Kelso (1980), parameters of movement other than location were given to participants in advance. People learned what direction of movement, extent of movement, and/or arm for movement were needed, but timing was affected only in conditions when precues were ambiguous because the input display had low stimulus-response compatibility. Under high stimulus-response compatibility, response times and MTs did not indicate that movement parameters are specified separately or serially.

From this research, it seems safe to assume Fitts’ Law holds in a 2-D framework by combining the multiple dimensions of movement into a single Euclidean distance. However, Jagacinski and Monk (1985) showed that movements from a central location along horizontal and vertical radii were faster than movements along diagonal radii. They rejected the Euclidean model of uniformly efficient movement space, meaning the Pythagorean theorem for combining height and width of motion did not provide a perfect solution. On the other hand, they also rejected the city-block model of movement in which total distance is the sum of both dimensions of movement. (Results of my experiment do not show this pattern of MTs. For the most common
distance of movement when participants indicated a category choice without confidence, mean MT for horizontal movement was 300 milliseconds and mean MT for diagonal movement was 288 milliseconds.)

Also in the original work of Fitts, “width” of a target was a single dimension running parallel to movement trajectory, but targets have other dimensions. While working to improve computer usability for people navigating a path through a 2-D graphical user interface of software, researchers developed a “steering law” as an extension of Fitts’ Law. This law has been independently discovered on more than one occasion, but the most comprehensive account of it was provided by Accot and Zhai (1997).

Assuming Fitts’ Law applies to the height of a target just as it does the width, researchers (e.g., Accot & Zhai, 1997) considered what happens when the primary need for a person’s movement is to stay within the boundaries of a path (height). As the space between the boundaries increases, the speed of movement increases. Imagine driving a car down a wide road versus a narrow road: it is safe to drive faster on the wide road than the narrow road. The same is true for moving a mouse pointer through computer application menus. If there is a specific target at the end of the path, the index of difficulty can be calculated and MT estimated by applying Fitts’ Law to a series of goals which are interposed between the start and end goal of movement. Fitts’ Law as applied to height is integrated along a continuous path where the goals are infinitesimal distances from each other or as width (as traditionally defined) goes to zero, $W \rightarrow 0$.

In a pilot study, I used the steering law to create the computer interface for responses and set high penalties for deviating from any one of the paths, but the way
participants performed did not seem to accord with with this law. Touchscreen areas were all of equal width but varying height. Height expansion was in proportion to distance of movement, but greater distances showed greater MTs when application of the steering law predicted that they should not. From this, I have come to believe that in this task people are not concerned with moving through particular areas to reach their goal areas. Instead, they take the most convenient route to reach goal areas. Because the joints responsible for reaching a finger to a target might more naturally move the finger in an arc, the two types of movement might lead to travel along divergent paths. If a person steers a finger down a bounded straight path toward a target, the movement should be straight, but if a person moves a finger efficiently and unboundedly toward a target, the movement could be curved as appropriate to the joints of the finger, wrist, or elbow.

By varying height but not width in the pilot study, the two dimensions of each area were never equal and may have created differential effects. Bohan, Longstaff, Van Gemmert, Rand, and Stelmach (2003) investigated whether the effects of height or width on MT were stronger. They found that with elliptical areas, decreases in MT were steeper for increases in target height than target width. Smaller width parallel to movement appeared to require less cautious movement than smaller height perpendicular to movement. When width was held constant, increasing height led to greatly decreasing MT.

In the pilot study, increased height was paired with increased distance so that MT could remain constant, but MT actually increased as well. A possible explanation comes from Murata (1999), who tested two models of 2-D movement. The “traditional” model treats $W$ in Fitts’ Law as the minimum of height and width.
His “effective” target width model creates a 95% ellipse on the 2-D joint PDF for endpoints of movement. The two models fitted about equally well, meaning that the smaller dimension played a strong role in determining MT. In my pilot study, the smaller dimension was almost always width. Although sensitivity to target height is greater than sensitivity to target width, Murata’s work suggests that it may be dampened by smaller target width.

For the present experiment, the areas in the interface have equal height and width to avoid and differential effects of the dimensions. I expected that enlarging touch areas in proportion to their distances would cause participants to move their fingers for nearly the same amount of time to make any confidence rating. Conversely, holding the size of touch areas constant should let movement times to different confidence ratings vary according to the distances of required movement. Importantly, what is true for MT should be true for MPT. As discussed in the introduction, Fitts’ Law that accounts for movement time may also account for additional time to begin movements.

3.2 Confidence Instructions

Another factor in the design is the demand for confidence judgments. In half of the conditions that I call confidence conditions, participants provide confidence ratings simultaneously with choice responses. In other words, participants’ confidence responses imply their choices (e.g., Egan, 1958). In the other half of the conditions, participants make no overt confidence judgments but must still touch specified areas when they make their choices. I refer to these as targeting conditions.
Figures 3.3 and 3.4 show targeting and confidence, respectively. In confidence conditions, participants point to areas that correspond to their subjective estimates of the probability of choosing the correct category, such that higher confidence ratings are a greater distance from the fixation point. In the targeting conditions, participants point to one of four areas which are all separated from the fixation point by the same distance randomly preselected by the computer.

I expect that having people pick their confidence ratings, and hence pick the distances their fingers must travel, will slightly increase their response times. In my model, times for moving from the fixation point to the particular touch areas should not change based on whether confidence ratings are required, but times to begin the movements should be greater when confidence ratings are required.

One explanation for why confidence is a part of the decision process that increases overall response time comes from the Hick-Hyman Law (Hick, 1952; Hyman, 1953), which resembles Fitts’ Law. It states that choice response time increases logarithmically with the number of alternatives. Using more than two confidence ratings increases the number of alternatives in a choice. Usher et al. (2002) proposed that in the context of dynamic choice models, the Hick-Hyman Law can be explained by increases to the decision threshold as the number of alternatives increases. In modeling the task for my experiment, the decision threshold is allowed to vary between confidence and targeting conditions to accommodate response time changes. Response time should change only because confidence ratings need more time to emerge from the decision process — and not because responding has become more complex in any other sense.
For some time, there was a controversy about the effects of task complexity on simple RT versus choice RT. Klapp, Wyatt, and Lingo (1974) showed that the complexity of a required response increased choice RT but did not reliably increase simple RT. Two decades later, Klapp (1995) resolved the question of when RT increases with task complexity. Using a Morse code analogue, he showed that choice RT depends on the duration of a response, and simple RT depends on the number of elements in a response (a series of actions versus a single action). Both aspects of responding change the “complexity” of the task, but in different ways.

In my experiment, the demand for confidence in some conditions could be viewed as adding an additional element to the response scheme. People must pick the right choice region and then the right confidence area within that region. Because I’m measuring choice RT, the work of Klapp suggests that an increase in RT should not appear for this type of added complexity. On the other hand, the level of confidence will determine the distance of movement and therefore the duration of responses, but any increases in RT should be accounted for by Fitts’ Law or Meyer’s OSM model.

### 3.3 Method for Experiment

The two factors, computer input display and need for confidence ratings, create a two-by-two, within-subjects design. People participated in a total of four conditions: rectangular regions for confidence, rectangular regions for targeting, triangular regions for confidence, and triangular regions for targeting. The rectangular confidence condition was expected to show a positive relation between confidence and MT which could be compared with the rectangular targeting condition that did not measure confidence. The triangular confidence condition was expected to show no relation
Figure 3.1: Distributions of stimuli. Probability densities of stimuli for each of four categories in the experiment.
between confidence and MT which could be compared with the triangular targeting condition that did not measure confidence.

### 3.3.1 Participants

Five participants were recruited from upper level psychology courses at The Ohio State University. All participants reported normal or corrected-to-normal vision as well as normal motor skills. Naive to the purpose of the experiments, they were tested individually after receiving both oral and written instructions. They were paid a minimum of six U.S. dollars per hour but had opportunities to earn more based on performance. Each person earned about $15 after participating in four sessions.

Recruiting and training the participants and having them complete four sessions required a large investment of effort and funding, but the payoff was expected to be cleaner data from dedicated people. The small sample size is partly justified by having each participant generate large quantities of data as is customary for fitting dynamic choice models.

### 3.3.2 Stimuli

On each trial of the experiment, a number from 1 through 99 appeared in the center of the screen. The number came from one of four categories. Each category was roughly described by a normal probability density function for stimuli. All categories shared a standard deviation of 13, but their means were separated by 20. The first category had a mean of 20, the second category had a mean of 40, the third category had a mean of 60, and the fourth category had a mean of 80. Because each stimulus needed to be between 0 and 100, the first and fourth categories were truncated on opposite outer tails so that they appear slightly skewed toward the middle of the
Figure 3.2: Ideal accuracy. Probability of a correct response to each stimulus when using the optimal choice strategy.

stimulus scale as can be seen in Figure 3.1. Any stimuli below one were re-sampled from the lower half of their originating distribution, and any stimuli above 99 were re-sampled from the upper half of their originating distribution.

When a stimulus closely fitted with one category, it typically did not fit with one other category. Rarely should it have been reasonable to participants to choose any/all of the four categories on a single trial because one, two, or three categories should have seemed unlikely. However, the variance of the category distributions ensured that no response strategy could result in only correct choices. The optimal observer in a signal detection framework would make correct choices on approximately
0.69 portion of trials. This strategy would involve choosing D when the number was less than 30, choosing S when the number was from 30 to 50, choosing T when the number was from 50 to 70, and choosing R when the number was greater than 70.

The accuracy for an optimal strategy would be 0.77 for the outer categories and 0.55 for the inner categories. Accuracy for the optimal strategy across the range of stimuli (over the course of infinite trials) can be seen in Figure 3.2. Choices become more difficult between the centers of the category distributions of stimuli (30, 50, & 70). This is because the probability that a stimulus in any of those areas comes from one category is lower in comparison to other categories. Instead, two categories were about equally likely for such stimuli.

This task lends itself to modeling of the response selection process because stimuli are quantified, and the probability distributions underlying the categories can be used to estimate the activation for a category in a sequential sampling paradigm (e.g., Ratcliff, Van Zandt, & McKoon, 1999). In modeling the participants’ responses with the BAC, the accumulation rate for a choice category during each trial of the experiment will on average equal the probability that the category generated the stimulus.

3.3.3 Apparatus

Timing of stimulus presentation was maintained by a Hewlett Packard XU7/500 desktop computer with Intel Pentium processor. The display of stimuli occurred on an attached 17-inch Elo Entuitive anti-glare CRT video monitor with resistive touchscreen capabilities (Actual diagonal measurement of the screen was 15.85 inches.). The monitor was mounted into a desk so that its screen was flush with the desk’s
surface. This rotation of the monitor avoided the common problem of fatigue in an unsupported arm extended toward a vertical screen for a prolonged period. 9.25 inches of length from the edge of the desk to the monitor allowed space for free movement of an elbow rest as a participant moved a finger over the screen.

The layout of choices that I used to achieve good stimulus-response mapping places shapes for choices radially about a fixation point (Similarly, a semicircle of buttons was used for confidence by Petrusic & Baranski, 2003.). This allowed each choice to be an equal distance from the starting location of movement. Returning to the fixation point each time should reduce response bias. (Augustyn & Rosenbaum, 2005, showed that people appear to have a basic understanding of Fitts’ Law which allows them to optimize their starting position when they expect to move to particular targets.) Because the categories of stimuli were equally spaced along the range of stimuli from low to high, choice regions for those categories were placed on the screen like spokes on a wheel with the lowest category at 180 degrees, the second category at 120 degrees, the third category at 60 degrees, and the highest category at 0 degrees.

The confidence conditions used a graphic rating scale for confidence judgments. The scale was based on distance of linear travel between the central pixel of the area where movement began and the pixel of a choice region where movement ended. The screen resolution was set to $640 \times 480$ pixels, so confidence could be considered on a continuous scale because distance was measured from 20 to more than 240 pixels. However, the pixel as a unit of measure is so small that the tip of a finger cannot reliably voluntarily travel any of these distances. For the ease of participants, the full scale was divided into 10 segments which allowed for some inaccuracy of movement.
Figure 3.3: Order of events for trials in the rectangular-regions-for-targeting condition of the experiment. When ready, a participant pressed an area near the center of the screen to see a stimulus. The person released pressure on the area after choosing and planning a response. Quickly, the person pressed the target area for that response and received feedback and earnings.
Figure 3.4: Order of events for trials in the triangular-regions-for-confidence condition of the experiment. When ready, a participant pressed an area near the center of the screen to see a stimulus. The person released pressure on the area after choosing and planning a response. Quickly, the person pressed the confidence level for that response and received feedback and earnings.
By treating confidence as subjective probability estimation, the full confidence scale was from 0 to 1, and dividing the scale into 10 levels of confidence made the centers of the levels correspond to probabilities of 0.05, 0.15, 0.25, 0.35 . . . 0.95. Because participants chose from four categories, chance performance was 0.25. Furthermore, because a pilot study showed that participants almost never used any level below the fourth (very few responses less than 0.30), the screen did not display the three lowest levels of confidence, leaving seven areas in a region to represent the levels of confidence in a choice.

The shapes of the regions differed between input display conditions. Two examples are shown in Figures 3.3 and 3.4. In the rectangular conditions, participants saw seven equal, square areas for the levels of confidence. The squares’ widths were all 22 pixels. In the triangular conditions, areas’ widths increased linearly with distance from the starting point. Participants saw seven trapezoidal areas increasing in average width from 10 pixels for the lowest confidence to 72 pixels for the highest confidence.

The major challenge for this computer interface was adjusting the areas of the targets so that each area had the same index of difficulty. A secondary challenge was keeping the seven remaining levels for confidence in the confines of the screen while making the smallest touch screen areas large enough to keep participants’ touch accuracy high. With the triangular regions, meeting this challenge required application of Fitts’ Law with some parameters held constant.

Levels of confidence, represented by touch areas, had to abut without overlap to maintain the graphic rating scale. Let \( d_i \) represent the distance from the center of the starting area to the center of a confidence area, and let \( w_i \) represent the width of the area at that distance \( d_i \). To simplify the problem, I assume each confidence area has
a height equal to its width despite not being square. The distance between \( d_{i-1} \) and \( d_i \) must be the sum of half of the height/width of each area, \( d_i - d_{i-1} = \frac{w_i}{2} + \frac{w_{i-1}}{2} \).

From Eq. (1.1), it can be shown that when the index of difficulty, \( I_d \), is constant, \( w_i = \frac{d_i}{2^{l_d-1}} \), so that \( \frac{w_i}{2} = \frac{d_i}{2^{l_d}} \). This means that

\[
d_i = d_{i-1} + \frac{d_{i-1} + d_i}{2^{l_d}}.
\]

By rearranging the above equation, one can see that \( d_i = d_{i-1} \left( \frac{2^{l_d+1}}{2^{l_d} - 1} \right) \), implying that a minimum distance, \( d_\alpha \), can serve as the basis for the distance of any area farther from the starting point, because

\[
d_i = d_{\alpha} \left( \frac{2^{l_d} + 1}{2^{l_d} - 1} \right)^i. \tag{3.1}
\]

The maximum distance, \( d_\omega \), is the far edge of the choice region. Because \( d_\omega = d_n + \frac{w_n}{2} \), where \( d_n \) is the farthest area, it is also the case that \( d_\omega = d_n \left( \frac{2^{l_d}+1}{2^{l_d} - 1} \right) \). Using Eq. (3.1) to find \( d_n \) yields

\[
d_\omega = \frac{d_\alpha (2^{l_d} + 1)^{n+1}}{2^{l_d}(2^{l_d} - 1)^n}. \tag{3.2}
\]

Fortunately, the first three areas of the regions were not used, so I was only concerned with \( d_\omega \) and \( w_4 \). From the previous equations, it can be shown that

\[
d_\omega = \frac{w_4 (2^{l_d} + 1)^{n-3}}{2(2^{l_d} - 1)^{n-4}}.
\]

Unfortunately, there is no easy way to solve for \( I_d \), so it is non-trivial to ensure that the shapes for the choice regions will have a total extent, \( d_\omega \), less than half the screen’s smaller dimension and also have their smallest target width, \( w_4 \), approximately the size of a person’s fingertip. Instead, one must assign a value to \( I_d \) which will satisfy these requirements. For a method generalizable to possible future experiments of this
nature, I set \( w_\alpha \) to 2 as a mathematical convenience, so Eq. (3.2) reduces to

\[
d_\omega = \frac{(d_\alpha + 1)^{n+1}}{(d_\alpha - 1)^n}.
\]  

(3.3)

I chose \( d_\alpha \) to make \( w_\alpha \) approximately 10 and \( d_\omega \) approximately 240. With my chosen \( d_\alpha \) and \( w_\alpha \), \( I_d \) was 2.48 for all areas of the triangular conditions. For the rectangular conditions, \( I \) ranged from 3.11 for the lowest confidence to 4.37 for the highest confidence.

There is some concern that Fitts’ Law breaks down for tasks with an index of difficulty less than two (e.g., Welford, 1968; Klapp, 1975) because the resulting MTs are so small (In the OSM, the maximum number of submovements for such short MTs is close to one.), but the problem has been avoided here.

3.3.4 Procedure

I obtained informed consent and provided instructions before the participants performed the task. Participants attended four sessions to provide data for the four conditions of the experiment. Rather than randomizing the order of conditions to counterbalance them, each participant experienced the conditions through one of four orderings. Controlling for the position of each condition in the sequence and first-order sequential effects, the reduced Latin square for the orderings is in Table 3.1. In this counterbalancing design, each condition was in each position once and was preceded by each other condition once.

Each session lasted 20 to 30 minutes. I required all sessions to be completed within a week of starting and to be separated from other sessions by a minimum of two hours.
Table 3.1: Latin square for the ordering of conditions. RT is rectangular targeting (not reaction time), RC is rectangular confidence, TT is triangular targeting, and TC is triangular confidence.

All participants in this experiment were told to pretend that the categories were climate regions of a country and the stimuli were inches of rainfall per year for different cities in that country. Their job was to pick the climate region to which each city belonged. Regions were initially labeled as D, dry grassland, S, savannah, T, tropical forest, and R, rain forest.

Data for each person in each condition were collected during four main sets of 50 trials each. For each trial, participants saw a black background with a white square containing a plus sign at the center of the screen. When they were ready, they pressed and held the area with an index finger to see a number near the center of the screen just above the area. At the same time, white choice regions appeared with divisions between seven levels. When participants decided which region had generated the stimulus and knew which level they would press to make a response, they released the area with the plus sign and began moving toward their target. The computer recorded the duration of each press and the time between release and re-press of the screen. Along with a statement of which region was correct for each trial, feedback after responses took the form of U.S. currency earned. Correct choices earned positive cents, and incorrect choices earned negative cents.
In the confidence conditions, participants were told that their movements would begin at the center, a place with no confidence, and end at a higher level of confidence, as high as complete certainty about a choice. The consequences of choices were commensurate with confidence: the lowest confidence level (0.35) added or subtracted four cents from participants’ earnings, and the highest confidence level (0.95) added or subtracted ten cents from their earnings. In the pilot study, the average confidence rating was approximately 0.70 for correct choices and approximately 0.55 for incorrect choices. This means people gained about 7.5 cents for correct choices and lost about 6 cents for incorrect choices.

In the targeting conditions, participants were instructed to touch predetermined levels which were unrelated to confidence. To keep all conditions as similar as possible aside from the intended manipulations, the preselected response locations for the targeting conditions were randomly drawn from distributions which approximate the empirically observed distributions of confidence judgments in the pilot study (See Figure 3.5.). If participants had gained or lost more cents for more distant predetermined targets, they might have been more careful about those choices as soon as a target level appeared. Instead, a participant making a correct choice (touching the correct region and level) gained seven cents, but a participant making a wrong choice (touching the incorrect region but correct level) lost six cents, regardless of which level was predetermined. Participants lost one half-cent extra for every level by which they missed a target, rounded up to the next whole number of cents.

In all conditions, participants lost eight cents for touching any black part of the display, and the screen told them that they touched outside the appropriate regions. If the duration of a press on the central area with the plus sign was less than 400
Figure 3.5: Histogram of predetermined target levels. In the targeting conditions, participants make a choice by pressing a randomly preselected level on one of four regions. Each target level is unrelated to confidence.
ms, the screen stated that the release of the area was too quick. If the latency of a re-press on the screen was more than 600 ms, the screen stated that the movement was too slow. These guidelines came from the distributions of RTs and MTs in the pilot study.

Each condition had three training phases. For all conditions, the first training phase was a set of 30 trials in which participants became familiar with the choice task by pressing regions for their choices. At the end, they saw the portion of their responses that were correct. The second training phase was a set of 30 trials in which regions were divided into levels. For targeting conditions, participants were told to touch a different level on each trial. For confidence conditions, participants were told to touch the level that corresponded with their confidence. At the end, they saw the portion of their responses that were correct and the average of the levels they had touched. For the confidence conditions, this feedback was meant to help participants calibrate their confidence with their accuracy. The third training phase was a set of 30 more difficult trials in which cents were given or taken on each trial in keeping with the condition. These trials resembled the trials to come after training. At the end, participants saw the portion of their responses that were correct, the average of the levels they touched, and the sum of the cents they earned.

Instructions that appeared on screen can be found in the appendix. Following each training phase and each of four sets of trials, participants had an opportunity to rest. At the end of each set consisting of 50 trials, they also received feedback about their performance. Participants saw the portion of their responses that were correct during the set, the average of the levels they touched during the set, and the sum of the cents they earned over all sets up to that time. Having 200 trials per
condition struck a balance between obtaining sufficient data for model-fitting and avoiding participant fatigue.

3.4 Results

Data from one participant were not included in reported results due to failure to follow instructions. In confidence conditions, the person used three levels of confidence almost exclusively. On many trials, his responses were outside the choice regions and therefore undefined. For other participants, these mistakes were rare and the trials on which they occurred were dropped from analysis.

Other data removed from analysis belonged to trials on which the time to begin the trial, the RT, or the MT were more than three standard deviations away from that participant’s mean or if response time was less than 300 ms.

3.4.1 Choice

The remaining participants chose each category with about the same frequency. Climate D was chosen on 0.23 portion of trials; climate S was chosen on 0.27 portion of trials; climate T was chosen on 0.26 portion of trials; and climate R was chosen on 0.23 portion of trials. The differences in average choice probabilities were primarily due to the first participant, P1, who chose the outer categories (D and R) only 0.40 portion of the time and the inner categories (S and T) 0.60 portion of the time.

The mean accuracy across trials of the experiment was 0.65. Mean accuracy was 0.78 for both of the outer categories and 0.53 for both of the inner categories. Across participants, accuracy ranged from 0.64 for participant P1 to 0.66 for participant P4.

Although the task was somewhat complicated, participants quickly achieved stable performance. Figure 3.6 shows the accuracy for different trials in the first condition.
Figure 3.6: Probability of a correct response as a function of the sequence of response. The plot combines accuracy information from all participants in the first condition in which each participated. The line represents a moving average across trials.
in which each person participated. There appears to be no improvement from start to finish, nor fatigue.

### 3.4.2 Confidence

The apparatus recorded the coordinates of the first pixel touched, so distances from the fixation point to the point of touch could be calculated geometrically. This offers one measure of confidence that could be used. Primarily, confidence was recorded on a discrete scale as 4 through 10 but rescaled to be between 0 and 1 as subjective probabilities. The lowest confidence was 0.35, and the highest confidence was 0.95.

Mean confidence over all choices ranged from 0.69 for P2 to 0.76 for P1. Examples of confidence distributions for each participant, labeled P1 through P4, are shown in Figure 3.7. Although each participant had a different response style, all showed a tendency to overuse one of the two highest confidence ratings. This is likely to have contributed to the general overconfidence in the experiment.

Mean confidence for choice D was 0.81, mean confidence for choice S was 0.66, mean confidence for choice T was 0.67, and mean confidence for choice R was 0.79.

Any extreme stimulus should provoke high confidence responses because only one category has a high likelihood of generating that stimulus. Similarly, the choice of that category should result in high accuracy. Thus, the correlation coefficient for confidence and accuracy across participants was 0.115, indicating a weak, positive linear relation.
3.4.3 Response Time and Reaction Time

Response times included both MTs and RTs. Some models were later judged by how well they predicted response times while others were judged by how well they predicted RTs. Analysis was performed primarily on RTs.

Response Time

In all cases, the computer recorded RT and MT as separate pieces of total response time. Times are reported in milliseconds (ms). Mean total response times ranged from 1323 ms for P4 to 1728 ms for P3.

For group data averaged across participants, mean total response time was 1389 ms for the rectangular targeting condition, 1586 ms for the rectangular confidence
condition, 1393 ms for the triangular targeting condition, and 1583 ms for the triangular confidence condition. It appears that confidence ratings require approximately an additional 200 ms over choices in the targeting condition. From this simple level of analysis, it also appears that the input display makes no difference in decision times, but later analyses revealed otherwise.

**Reaction Time**

RTs are a better approximation of decision times than response times are. The range of RT means across participants was from 986 ms for P4 to 1394 ms for P3.

From the model, I predicted that response times would be greater for confidence conditions. Furthermore they would decrease as confidence and accuracy increased, especially in the triangular confidence condition where movement would not impact RT differentially. In the rectangular targeting condition, the model showed that reduced RT would increase as target distance increased, but in the triangular targeting condition they would be the same for varying target distances.

Figure 3.8 shows the distributions and means of RT collapsed across participants for each condition. They are Vincent histograms (See Van Zandt, 2000a.) whereby quantiles of participants are averaged. Some of the difference between total response time in confidence and targeting conditions has been removed with the subtraction of MTs, and where no difference between rectangular and triangular conditions appeared for response time, there is about a 30 ms difference for RT. Confidence takes much longer than targeting, and triangular responding takes slightly longer than rectangular responding.

To see how RT relates to confidence and accuracy, I graphed their means across the values of the stimuli after pooling all participants’ data (to be sure all stimuli
Figure 3.8: Reaction time distributions. The plots show Vincent histograms of RT for group data from all four conditions of the experiment.
Figure 3.9: Three decision measurements for the range of stimuli. Accuracy, confidence, and RT means are shown as a smoothed function of stimulus in confidence conditions. Accuracy, target level, and RT means are shown as a smoothed function of stimulus in targeting conditions.
had enough data points). Figure 3.9 shows these means with a standard smoothing algorithm applied to them to make the patterns more obvious. This uses Tukey’s (1977) method of calculating running averages over three points until convergence. The lines for accuracy resemble the one for the optimal strategy in Figure 3.2. The relation between accuracy and RT is most clear in the triangular targeting condition where they generally move in opposite directions in response to stimuli. The relation between confidence and RT, as well as the relation between confidence and accuracy, is most clear in the rectangular confidence condition.

In confidence conditions, RT clearly has a negative relation with confidence \( r = -0.315 \). As expected in targeting conditions, RT doesn’t have any relation to target level. In all conditions, RT has a negative relation with accuracy as well \( r = -0.104 \). The pattern for this is less apparent because accuracy is more volatile (It is measured as 0 or 1.) than confidence. Again, confidence and accuracy have a positive relation.

The only major deviation from the data patterns for the group was for P3, who had a stronger relation between RT and confidence \( r = -0.574 \).

For a better understanding of how the full distributions of confidence/target level and RT relate to each other, the plots in Figure 3.10 show the seven levels of confidence or targeting and five quantiles of RT averaged across participants. Again, RTs are higher in confidence conditions than they are in targeting conditions.

Also notice that in the rectangular confidence condition, as confidence increases, not only the median of RT decreases but also the spread of the RT quantiles decreases. These patterns are less clear for the triangular confidence condition, though I had expected the downward trend to be stronger than in the rectangular confidence condition because of motor programming differences.
Figure 3.10 shows that for targeting conditions, RT distributions have similar spreads across the range of target distances. RT quantiles in targeting conditions are nearly flat, which the model did not predict. Indeed, the correlation between target level and RT is -0.020 for both rectangular and triangular conditions. This suggests that MPT bears little relation to MT, because in the following section, I will show that MT varies systematically with target distance in the rectangular targeting condition while RT does not. However, I will also show that MT and RT share different relations in rectangular and triangular conditions, suggesting that the triangular conditions reduced the impact of movement on RT.

### 3.4.4 Movement Time

Mean MT over all conditions was extremely similar for all but one participant. Participants 2, 3, and 4 had a mean MT of 335 ms, but participant 1 had a mean MT of 209 ms.

For calculating the index of difficulty, Euclidean distance to the center of each touch area was used because average MTs for equal distances of movement in different directions were generally similar. For example, for confidence conditions, the most common confidence response had a group mean MT of 338 ms for choice D, 335 ms for choice S, 330 ms for choice T, and 339 ms for choice R.

Although confidence conditions tie the properties of touchscreen areas to confidence, my model assumes that MT is not directly related to confidence. The full model incorporates the OSM model, which predicts that MTs increase with the index of difficulty, \((D/W)^{1/n}\), so MTs should increase with target distance in rectangular conditions because target widths are all the same. However, MTs should not change
Figure 3.10: RT quantiles for levels of confidence/targeting. The different symbols represent the quantiles of the RT distributions averaged over participants. The numbers at the top are the 0.9 quantiles, where the number indicates the confidence or target level of responses; open triangles are the 0.7 quantiles; filled diamonds are the 0.5 quantiles; open squares are the 0.3 quantiles; and open circles are the 0.1 quantiles.
with target distance in triangular conditions because target widths increase with target distance to maintain the index of difficulty.

MTs were not fully as expected based on previous research. Providing confidence ratings seems connected to a slowing of MTs on both input displays. The mean MT across participants was 300 ms in the rectangular targeting condition and 339 ms in the rectangular confidence condition ($t = 8.28, df = 1473, p < 0.05$). The difference was significant and of medium size. Cohen’s $d$ was 0.603. The mean MT across participants was 284 ms in the triangular targeting condition and 296 ms in the triangular confidence condition ($t = 2.79, df = 1407, p < 0.05$). Although the difference was significant, the statistical test appears to be driven higher by the large sample size. Cohen’s $d$ was 0.211, indicating a small effect in the triangular condition. These tests were not planned in advance but suggest that the dependency between MT and RT is bidirectional. When a confidence rating takes longer than a choice, the MT for the response increases. Furthermore, when MT is longer for a response, the choice or confidence RT increases. The focus of this research is on the latter effect, although the former effect is worth exploring in future research.

Figure 3.11 was constructed just as Figure 3.10 but shows the quantiles of MTs broken down by levels of confidence or targeting, depending on condition. The scale along the $x$-axis is maintained on the plots by using the relative distance of targets in the left two plots.

MT as a function of target distance was just as expected for the rectangular targeting condition. MTs increased as distance of movement increased. Some of the MTs in the triangular targeting condition were not as expected. MTs were fairly flat across the range of target distance where target width was chosen to maintain
Figure 3.11: MT quantiles for levels of confidence/targeting. The different symbols represent the quantiles of the MT distributions averaged over participants. The numbers at the top are the 0.9 quantiles, where the number indicates the confidence or target level of responses; open triangles are the 0.7 quantiles; filled diamonds are the 0.5 quantiles; open squares are the 0.3 quantiles; and open circles are the 0.1 quantiles.
the index of difficulty. However, variability was smaller for smaller distances. Although the variance and mean of many distributions are positively related, MTs have historically been fitted with least-squares regressions for which homoskedasticity is assumed. Furthermore, MTs were on average longer for a few of the nearer targets.

Another surprise in the MTs is where they fell for the highest and lowest confidence in the triangular confidence condition. MTs across the other levels of confidence were fairly flat, but the lowest confidence had the largest MTs, and the highest confidence had the smallest MTs.

The peculiarities of the MT data force me to acknowledge that the triangular input display might not have been completely successful at equalizing MTs across target levels. A fair amount of variability appeared for MTs when people used the triangular input display. Some of this variability might be accounted for by considering the “effective” index of difficulty for movements. Welford (1968) suggested that in some cases people might not be performing the movement task as a researcher intends and recommended calculating the index of difficulty based on the actual endpoints of movement. Often, the narrower targets in the triangular targeting condition had greater MTs even though they were nearer the starting point of movement. The correlation coefficient for MT and target distance was \(-0.109\) in this condition. It was possible that the average movement to farther targets fell short of the centers of the targets or that the range of movement endpoints for closer targets was much less than the true width of the target, hence the effective index of difficulty might decrease with increasing target level although nominal index of difficulty was constant. This issue will be addressed more thoroughly with model fitting in the following chapter.
The decreasing MTs for increasing movement distances also appears for the triangular confidence condition. At first glance the relation seems stronger, but notice that for five of seven confidence ratings, MT is relatively stable. The highest and lowest confidence drive the perception of a pattern. The correlation coefficient is -0.244 until the highest confidence ratings are removed from consideration, then it goes to -0.005. This is not entirely unexpected: Bradi, Adam, Fischer, and Pratt (2009) found that when targets are arranged in an array, the farthest target is special. In an apparent violation of Fitts’ Law, the farthest target may be reached faster than the second and third farthest targets. However, this is an inadequate explanation because people reached the farthest target faster than any other in the triangular confidence condition but slower than any other in the rectangular targeting condition. The farthest targets in the input array should have consistently shown deviation from Fitts’ Law, but only did in half of the conditions.

The MTs in the rectangular confidence condition show a pattern like those in the rectangular targeting condition but with added concavity. There is a general positive trend combined with an inverted U-shape. As in the triangular confidence condition, participants moved to the highest confidence level more quickly than several other confidence levels.

MT had no obvious connection to accuracy for any condition or participant. However, MT and RT were related. Earlier, Figure 3.10 suggested that any MPT included in RT did not follow the same trends as MT because RT was not affected by the same variables as MT; however, for the rectangular input display, MT and RT had a positive correlation ($r = 0.216$). With the full model, I predicted that MTs would increase for longer movements in the rectangular conditions and influence RTs through MPTs. I
also predicted that by adjusting touch screen areas in triangular conditions according to Fitts’ Law, MTs would not be affected by movement distance and therefore would not influence RTs. Indeed, for the triangular input display, MT and RT had a very weak correlation ($r = 0.042$).

### 3.5 Discussion

For relations among accuracy, confidence, and RT, my expectations were correct: accuracy and confidence were inversely related to RT. However, for the more interesting model predictions regarding MT and RT, the experimental results make it difficult to draw firm conclusions. MTs correlated with RTs in the rectangular conditions, but perhaps not for the reasons I had imagined. There, the MTs increased with increasing target distance, but RTs did not. MTs did not correlate with RTs in the triangular conditions. There, MTs unexpectedly decreased with increasing target distance, but RTs did not.

Figure 3.12 should be compared to Figure 3.13. The first graph displays the anticipated relations among confidence, MT, RT, and response time in the four conditions of the experiment. The second graph displays the results obtained from the experiment. They are not too different from each other. The proposed model has captured the general trends in the data. Assuming that MPT was part of RT was important for generating the correct ordering of mean RTs across input displays for the highest confidence/target levels. As predicted, triangular conditions solicited faster reactions from participants than rectangular conditions as the target or confidence level increased.
Figure 3.12: Predicted results for MT and RT. The plot on the left shows the trend for MT (lower lines) and RT (upper lines), while the plot on the right shows the total response time when MT and RT are summed.
Figure 3.13: Empirical results for MT and RT. The plot on the left shows the mean MT (lower lines) and RT (upper lines) as smoothed functions of confidence/target level, while the plot on the right shows the response time when MT and RT are summed.
The obvious deviation of the data from the model predictions is where RTs are large for some of the middle confidence in the triangular confidence condition. The predictions in Figure 3.12 are grossly simplified to show the general patterns expected for reasonable parameter values of the model. The curve in the real pattern of results for the triangular confidence condition can be explained through the model by changing the frequencies of different confidence responses. If people prefer not to use the low confidence responses, decreases in RT with increases in confidence only become noticeable for half of the confidence scale. Thus, average RT is fairly flat across the lowest levels of confidence in the results from the triangular condition. In the next chapter, I show how the model can adapt somewhat to non-monotonic trends such as this. Figure 4.10 shows that the model best fits the results of the triangular confidence condition by producing great change in RTs for confidence levels six through nine and very little change for levels four through six.

The many places where the two graphs match offer weak support to my theory regarding the dependency between the movement and response preparation processes in the decision task, but the model evaluation in the next chapter will more conclusively show which theoretical description of the relations among variables is best.
Chapter 4: MODEL EVALUATION

The theory behind this work is expressed in variations on the BAC model. My expectation was that the full model, which accounts for MT with the OSM and takes one of those parameters for use in accounting for MPT, will be able to reproduce the RTs in the experiment better than alternative versions of the model. Furthermore, I expected that modeling of total response time would be better with data from triangular conditions than rectangular conditions because the triangular condition’s input display was constructed to equalize MTs.

4.1 Models

There are six model fits based on the BAC and five essential comparisons among them. Distinctions between the model fits can be seen in Table 4.1. The basic model adds a constant to decision time in computing response time or RT. This may be $\alpha$ to describe all non-decision times in the task or $\alpha_p$ to describe non-decision times that contribute only to RT. The expanded model has $\alpha_p$ but also incorporates a form of the optimized-submovement model to account for MTs, MPTs, or both.

Models are fitted to probabilities of correct and incorrect choices and confidence. Each model is also fitted to either total response time or RT for one confidence condition and one targeting condition within an input display type. Fitting the
models to the targeting task reveals whether consideration for MPT improves dynamic choice models. Fitting the models to the confidence task reveals whether the models can handle additional variables and their interactions. For simplicity, all models assume that the areas touched were those intended by the participants.

The basic model applied to response times in rectangular conditions (BasicRespR) is the standard style of dynamic confidence model. It attempts to account for response times when confidence ratings are provided in the fashion of the rectangular input display. The basic model for response times in triangular conditions (BasicRespT) has all the same parameters as BasicRespR but is fitted to the triangular conditions.

The basic model applied to RTs in the rectangular conditions (BasicReacR) has the $\alpha$ parameter from the OSM, but it no longer accounts for MT, but still possibly accounts for MPT, so it receives a subscript to become $\alpha_p$.
The expanded model is fitted to RTs in rectangular conditions by using two additional parameters, $\beta_p$ and $n_p$, to account for MPTs using the structure of the OSM. In the first version (FProgReacR), $n_p$, the maximum number of submovements, is a free parameter. In the second version (CProgReacR), $n_p$ is constrained. The OSM is first employed for explaining MTs, and the parameter $n$ is replicated as $n_p$ in the expanded model. The third version also with $n_p$ constrained is fitted to RTs from triangular conditions (CProgReacT).

4.2 Model Selection

Because the models vary in complexity, a model selection criterion one employs should penalize greater complexity. When fitting models, it is possible to achieve better fits by adding more parameters or making the form more flexible (Myung, 2000). When the set of parameters which achieve the best fit do not generalize to other data sets, this is called “overfitting”. In the absence of new data, the model may appear better than it really is.

A model selection criterion which was designed to penalize model complexity but primarily penalizes the number of free parameters is the Bayesian Information Criterion (BIC). The BIC (Schwarz, 1978) is based on the likelihood function of a model, which I do not use for these models. In the typical equation,

$$BIC = -2 \log(L) + M \log(N),$$

where $M$ is the number of free parameters, $N$ is the number of data points, and $L$ is the likelihood of the data for the model. A larger likelihood, or its approximation, indicates that the model is better able to produce the data observed. By including the negative likelihood, lower values of the BIC indicate better model fits.
Simulation of a model allows approximation of its likelihood function. For this project, the likelihood can be approximated by binning the distributions of choices, confidence ratings, RTs, response times, and MTs. The number of simulated data points falling into those bins can be used to estimate the probability that real data should fall into those bins if the model were correct. The reformulation of the BIC for binned data is

\[
BIC = -2 \left( \sum n_i \log(\pi_i) \right) + M \log(N),
\]

(4.1)

where \( n_i \) is the number of observations in bin \( i \), and \( \pi_i \) is the portion of data predicted by the model to be in bin \( i \). The left terms of the two formulas are similar. For a model that fits well, differences in these two forms of the BIC will be negligible. The right terms are the same and penalize each model in proportion to its number of free parameters and the size of the sample to which it’s fitted.

### 4.2.1 Model Fitting

In fitting the models, a separate statistic is used which can also make use of binned data: the \( \chi^2 \) statistic. It compares the number of observed items in a bin to the number of expected items in a bin:

\[
\chi^2 = \sum \frac{(n_i - N\pi_i)^2}{N\pi_i}.
\]

(4.2)

During model fitting, an optimization algorithm attempts to minimize the \( \chi^2 \) over the bins for a model by finely tuning its free parameters. Any response time or RT distribution will be divided into five roughly equal parts, and confidence is already discretized into seven levels. For fitting the choice models, accuracy has two levels, so each condition had 70 bins (\( 5 \times 7 \times 2 = 70 \)). For fitting a movement time model, each condition had 35 bins because choice accuracy was not of interest.
Several optimization routines were tried on data from the pilot study, and the most successful at minimizing the $\chi^2$ statistic was the Nelder-Mead simplex algorithm (Nelder & Mead, 1965). A simplex algorithm is a routine that searches for a global minimum by constructing a simplex, an $M$-dimensional polygon with $M + 1$ vertices in the space created by $M$ parameters. At each vertex, the fit (or lack-of-fit) statistic is computed. The simplex is rotated around the vertex with the smallest value. These steps are repeated until the simplex shrinks to a sufficiently small size or a maximum number of repetitions is reached.

Although the simplex is generally robust to poorly behaved models (such as one with noisy simulations) and poor starting values of parameters, it still may fail to find the global minimum for $\chi^2$. To improve performance of the algorithm, I chose two reasonable starting values for each free parameter. For the dynamic choice/confidence models, I fitted the parameters in small batches after fitting them all simultaneously, then repeated the process.

All models were fitted to participants individually, and their BICs were computed on predictions for all participants using individual sets of parameters. Critical parameter estimates and fit statistics are reported in Table 4.1. All parameter estimates can be found in Tables A.1 and A.2 of the appendix.

### 4.2.2 Movement Time Models

Before fitting the dynamic choice/confidence models to the data, I first checked the quality of the OSM for fitting the MT data. The OSM has in common with Fitts’ Law the $\alpha$ and $\beta$ parameters but carries the additional $n$ parameter. Because the value of the $n$ parameter will feed into the MPT portion of the expanded BAC model,
I should justify the use of the parameter even for MT. Furthermore, any model must adequately fit the MTs before incorporating that model into a related choice model.

Although the OSM is popular and has more flexibility than Fitts’ Law to adapt to the behavior of individual participants, Fitts’ Law has been highly regarded for many decades and offers parsimony to the theory. Both models were fitted to the MT data from the experiment to see which performed better.

Because the graphs of the MTs for the four conditions of the experiment were not as expected from previous MT research, I also fitted the models using “effective” index of difficulty. The effective index of difficulty uses the spatial results of movement to find the average distance of movement to a target and the spread of endpoints of movement to a target (Welford, 1968; Murata, 1999). Essentially, the distance and width in the models are calculated from results in cases when people do not perform a task as researchers intend. People may not move to the centers of targets or may cluster their responses in a small part of the target so that nominal target distance and width are inappropriate for fitting the models.

Effective index of difficulty in this study uses the known methods to find the mean of the 2-D distribution of movement endpoints for calculating effective target distance and also creates a 95% confidence interval (3.92 standard deviations around the mean) of the endpoints for calculating effective target width. See Murata (1999) for further details.

In all, there were four models of MT: Fitts’ Law with nominal index, Fitts’ Law with effective index, the OSM with nominal index, and the OSM with effective index. The qualitative fits of the models can be seen in Figures 4.1, 4.2, 4.3, and 4.4.
Figure 4.1: Fitt’s Law using nominal index of difficulty to predict movement time quantiles for levels of confidence/targeting. The different symbols represent the quantiles of the RT distributions averaged over participants. The numbers at the top are the 0.9 quantiles, where the number indicates the confidence or target level of responses; open triangles are the 0.7 quantiles; filled diamonds are the 0.5 quantiles; open squares are the 0.3 quantiles; and open circles are the 0.1 quantiles. Lines represent the predictions of the model.
Figure 4.2: Fitts’ Law using effective index of difficulty to predict movement time quantiles for levels of confidence/targeting. The different symbols represent the quantiles of the RT distributions averaged over participants. The numbers at the top are the 0.9 quantiles, where the number indicates the confidence or target level of responses; open triangles are the 0.7 quantiles; filled diamonds are the 0.5 quantiles; open squares are the 0.3 quantiles; and open circles are the 0.1 quantiles. Lines represent the predictions of the model.
Figure 4.3: The optimized-submovement model using nominal index of difficulty to predict movement time quantiles for levels of confidence/targeting. The different symbols represent the quantiles of the RT distributions averaged over participants. The numbers at the top are the 0.9 quantiles, where the number indicates the confidence or target level of responses; open triangles are the 0.7 quantiles; filled diamonds are the 0.5 quantiles; open squares are the 0.3 quantiles; and open circles are the 0.1 quantiles. Lines represent the predictions of the model.
Figure 4.4: The optimized-submovement model using effective index of difficulty to predict movement time quantiles for levels of confidence/targeting. The different symbols represent the quantiles of the RT distributions averaged over participants. The numbers at the top are the 0.9 quantiles, where the number indicates the confidence or target level of responses; open triangles are the 0.7 quantiles; filled diamonds are the 0.5 quantiles; open squares are the 0.3 quantiles; and open circles are the 0.1 quantiles. Lines represent the predictions of the model.
One can see that the best fit is for the OSM using nominal index of difficulty. First notice that it does something the other three models do as well. It accommodates deviations from the straightforward predictions of MT models. The kinks in data patterns of any condition are closely followed by the predictions of the models. The curve in MTs from the rectangular confidence condition is fitted somewhat by all models. These patterns are explained through changes to the values of parameters for individuals in different conditions. It appears that the behaviors of participants were sometimes peculiar to the sessions in which they performed. The idiosyncrasies of the individuals and inconsistency of performance from day to day moved the MT distributions in unexpected ways that are well accounted for by the models.

How the nominal OSM fitted better than the effective OSM and effective Fitts’ Law is in the rectangular conditions where it more easily captured the general positive trend in quantiles across levels of confidence/targeting. Typically the effective index of difficulty improves model fit when nominal index of difficulty does not explain the data well. For this experiment, my concern that nominal index of difficulty would not sufficiently account for the variability in MTs was wrong. Effective index of difficulty did not increase as rapidly as nominal index of difficulty when level of confidence/target increased. Primarily this was attributed to effective target distances smaller than nominal target distances for higher confidence/target levels combined with minimal deviations between effective target width and nominal target width. As people reached to targets that were farther away from the starting point, they tended to fall short of the centers of the targets. Effective target distance did not increase as quickly as nominal target distance so that effective index of difficulty did not increase as quickly as nominal index of difficulty.
How the nominal OSM fitted better than nominal Fitt’s Law is less obvious from the graphs than the other contrasts between models. The qualitative fits of the two models are not too different, but the fit statistics are different. Table A.2 shows the $\chi^2$ values of the models. Nominal OSM has a fit statistic of 274, while nominal Fitts’ Law has a fit statistic of 298. Although both statistics indicated a significant difference between the model predictions and the data, the fit is considerably better for the nominal OSM. Because the OSM has an additional free parameter, we should consult the BIC which penalizes models for their numbers of parameters. The BIC comes out in favor of the nominal OSM as well.

From a theoretical perspective, Fitts’ Law holds intuitive appeal and makes a reasonable choice for modeling MTs and MPTs, but the statistical results suggest that the OSM is quantitatively superior. In trying to relate MPT to MT, the extra free parameter $n$ in the OSM offers a way to connect modeling efforts for both variables and also improves overall model fit.

4.2.3 Dynamic Choice/Confidence Models

The core simulation for all dynamic choice/confidence model fittings randomly samples $a$, the starting point, from the appropriate uniform distribution and four $d_i$, the accumulation rates, from their appropriate exponential distributions. Each exponential distribution is controlled by rate $v_i$, but this is not a free parameter for any model. Each $v_i$ is the inverse ratio of the probability that the stimulus in a trial was generated from the $i$th category to the probability that the stimulus was generated by any category. It is the reciprocal of a simple likelihood ratio which will
produce an average $d_i$ equal to the likelihood ratio. This gives no model a special advantage in predicting choices.

The accumulation rates must sum to a constant in an implementation of processing capacity that approximates the Hick-Hyman Law (See Brown & Heathcote, 2005.). If the number of choice alternatives varied in the experiment, this feature of the models would be more important in fitting the data. For my purpose, it also changes the predicted relation between confidence and RT. The highest confidence ratings in the experiment were to extreme stimuli, which had a very low chance of appearing in the trials but were strongly indicative of a particular category. The lowest confidence ratings were for the most common stimuli which could have come from multiple categories. Adjusting each $d_i$ lets DT decrease when only one alternative seems reasonable but overall activation of accumulators is limited. That is, the accumulators speed up when stimuli are closely tied to a single category but are rare.

The time to complete the core choice/confidence, $t_d$, is calculated as the minimum time for any accumulator to reach threshold $b$. Confidence is computed as $c_i(t) = z_i(t) / \sum_{j=1}^{N} z_j(t)$ from the evidence on the accumulators, and evidence for each accumulator is simply $z_i = a + d_i(t_d)$. This is confidence on a continuous scale, but participants are using a discrete scale, so the seven ratings are separated by six confidence criteria, $c_4, c_5, ... c_9$, placed between zero and one.

For the basic model, nine free parameters are needed for each person in each confidence condition: $A, b, c_4, c_5, c_6, c_7, c_8, c_9$, and $\alpha$ or $\alpha_p$. These are respectively the range of the starting point distributions, the decision threshold, the six confidence criteria, and the constant non-decision time or MPT. For targeting conditions, the confidence criteria are unnecessary, so only three free parameters are needed. For
each of the fittings of the basic model, 12 free parameters per person, or 48 across participants, were used because fittings were to one targeting and one confidence condition.

For expanded models, 11 free parameters are needed per person in each confidence condition because along with the first nine, $\beta_p$ and $n_p$ enter to account for MPT. In the targeting conditions, only five free parameters are needed for each person. Across participants, 64 free parameters were used for the FProgReacR fitting. For CProgReacR and CProgReacT, $n_p$ is not a free parameter, so two fewer free parameters were needed per person. On those two fittings, 56 free parameters were used in total.

**Qualitative Model Fits**

The qualitative fits of the models can be seen in graphs of the models’ predictions. Take Figure 4.5 as an example: many of the basic model’s predicted response time quantiles in the BasicRespR fit are too low. In the targeting condition, this is most apparent for the lowest quantiles and the highest target level. Conversely, underpredictions of response times are less common for the highest confidence ratings, which likely have great influence on the parameters of the model.

Throughout the model fits pictured, the points for low confidence/targeting are often not matched by the lines for the model predictions, but readers should not linger on these areas of the graphs because they are the most unreliable points. The lower levels of confidence and targeting were the least used for responding by participants, so they have the highest error associated with them. Furthermore, so few data were associated with one of the three lowest confidence ratings for each participant in at least one condition, that for the purpose of fitting the data, parts of the array had to be collapsed. The $\chi^2$ statistic performs poorly when a cell count is very small,
Figure 4.5: Predictions from the basic model of response time quantiles for levels of confidence/targeting in rectangular conditions. The different symbols represent the quantiles of the RT distributions averaged over participants. The numbers at the top are the 0.9 quantiles, where the number indicates the confidence or target level of responses; open triangles are the 0.7 quantiles; filled diamonds are the 0.5 quantiles; open squares are the 0.3 quantiles; and open circles are the 0.1 quantiles. Lines represent the predictions of model fit BasicRespR.
Figure 4.6: Predictions from the basic model of response time quantiles for levels of confidence/targeting in triangular conditions. The different symbols represent the quantiles of the RT distributions averaged over participants. The numbers at the top are the 0.9 quantiles, where the number indicates the confidence or target level of responses; open triangles are the 0.7 quantiles; filled diamonds are the 0.5 quantiles; open squares are the 0.3 quantiles; and open circles are the 0.1 quantiles. Lines represent the predictions of model fit BasicRespT.
Figure 4.7: Predictions from the basic model of RT quantiles for levels of confidence/targeting in rectangular conditions. The different symbols represent the quantiles of the RT distributions averaged over participants. The numbers at the top are the 0.9 quantiles, where the number indicates the confidence or target level of responses; open triangles are the 0.7 quantiles; filled diamonds are the 0.5 quantiles; open squares are the 0.3 quantiles; and open circles are the 0.1 quantiles. Lines represent the predictions of model fit BasicReacR.
Figure 4.8: Predictions from the expanded model of RT quantiles for levels of confidence/targeting in rectangular conditions. The different symbols represent the quantiles of the RT distributions averaged over participants. The numbers at the top are the 0.9 quantiles, where the number indicates the confidence or target level of responses; open triangles are the 0.7 quantiles; filled diamonds are the 0.5 quantiles; open squares are the 0.3 quantiles; and open circles are the 0.1 quantiles. Lines represent the predictions of model fit FProgReacR.
Figure 4.9: Predictions from the expanded model with constraint on \( n_p \) of RT quantiles for levels of confidence/targeting in rectangular conditions. The different symbols represent the quantiles of the RT distributions averaged over participants. The numbers at the top are the 0.9 quantiles, where the number indicates the confidence or target level of responses; open triangles are the 0.7 quantiles; filled diamonds are the 0.5 quantiles; open squares are the 0.3 quantiles; and open circles are the 0.1 quantiles. Lines represent the predictions of model fit CProgReacR.
Figure 4.10: Predictions from the expanded model with constraint on $n_p$ of RT quantiles for levels of confidence/targeting in triangular conditions. The different symbols represent the quantiles of the RT distributions averaged over participants. The numbers at the top are the 0.9 quantiles, where the number indicates the confidence or target level of responses; open triangles are the 0.7 quantiles; filled diamonds are the 0.5 quantiles; open squares are the 0.3 quantiles; and open circles are the 0.1 quantiles. Lines represent the predictions of model fit CProgReacT.
so cells of different confidence were combined to raise the number of observations in some parts of the array. This means that the models often were not able to match the data patterns for confidence ratings four or five.

Where the BasicRespR fit fails is in capturing the slight upward tilt of the response times in the rectangular targeting condition and the slight hill in response times in the rectangular confidence condition. Because these data patterns are exclusive to the rectangular conditions, the BasicRespT fit of the basic model to response times in triangular conditions in Figure 4.6 does not show the same problems. The fit to the triangular targeting data is very good with some underprediction of the highest quantiles. The fit to the triangular confidence condition is good for confidence ratings six through nine, but poor for confidence rating 10.

Notice that the response time patterns look somewhat like the MT patterns in Figure 3.11 because MT is part of response time. Because we know that the basic model adds a constant to response time for MTs that are not at all constant, the lack of fit is no surprise. The one condition in which the model performs well on response times is triangular targeting where MTs are nearly flat.

The third model fit, BasicReacR in Figure 4.7, appears to be qualitatively as good as BasicRespT. The predictions for the rectangular condition are closer to the data than before because RTs rather than response times were fitted. The influence of movement appears to be smaller for RTs, and the model’s lack of accounting for it is less detrimental here.

Predictions for RT in the rectangular conditions from different models are shown in Figures 4.7, 4.8, and 4.9 look similar to each other. However, Table 4.1 shows
that the fit statistics differ. How the values of the parameters and fit statistics vary between conditions and participants can be seen in Table A.1 in the appendix.

**Quantitative Model Fits**

If one considers the $\chi^2$ values in Table 4.1 as goodness-of-fit measures to reject or retain each model, all these models would be rejected. A known problem with the statistic in model fitting is that it increases as the size of the sample or array increases. To find a standard of comparison, I simulated the basic model for response times in rectangular conditions (like BasicRespR) for participant 2 and fitted it back to itself. When the optimization almost perfectly recovered the parameters of the simulation, the $\chi^2$ was 139, which would also result in rejection of the model.

The $\chi^2$ statistic is typically used to reject a model when the amount of mismatch between predictions and data is statistically significant. Ideally, the number of predictions and data in each bin would be perfectly equal. A perfect fit would render $\chi^2$ equal to zero. What would $\chi^2$ be for the worst fit? This is a difficult question to answer because both the numerator and the denominator of the equation change based on the predictions of the model. Knowing how the fit statistic relates to the best and worst possible cases would be helpful, so we can define a new statistic as $\chi$, where

$$
\chi = \sum |O_i - E_i|.
$$

This equation simply sums the number of predictions that do not fit data in each bin. The best fit would have a $\chi$ of zero, and the worst fit would have a $\chi$ equal to the number of observations. With these models, the predictions could fall completely outside the range of the data so that no observations would be in the bins where
predictions would be. In the first model fit, BasicRespR, the \( \chi \) was 585 compared to a maximum of 1475, indicating that 585 of 1475 simulated data were in the wrong bins. In the second model fit, BasicRespT, the \( \chi \) was 520 compared to a maximum of 1409, indicating that 520 of 1409 simulated data were in the wrong bins. By converting these to percentages, the BasicRespR fitting correctly binned 60.3\% of simulations, and the BasicRespT fitting correctly binned 63.1\% of simulations. The BasicReacR fitting correctly binned 64.4\%, the FProgReacR fitting correctly binned 65.5\%, the CProgReacR fitting correctly binned 65.5\%, and the CProgReacT fitting correctly binned 65.7\%. Of course, these estimates can be conservative because misplaced predictions will be counted twice when missing from the correct bin and present in the wrong bin. If these values were the basis for model evaluation, the relative value of each would still let me choose which model is best, but I’m primarily using the BIC to compare models instead.

4.2.4 Model Comparison

The first essential comparison of model fits is between BasicRespR and BasicRespT. The BIC and \( \chi^2 \) are smaller for the latter, suggesting that the basic model fits data from the triangular conditions better than the rectangular conditions.

The second essential comparison of model fits is between BasicRespR and BasicReacR. This is like a test of whether RT improves model fit under normal circumstances. It appears that with the traditional confidence input display, the predictions of the model are closer to RT than total response time. Together, these support the idea that MPT is smaller than MT, and both are unaccounted for by the basic model.
A third comparison can be made between BasicReacR and FProgReacR. The latter is the expanded model, which attempts to account for MPTs within RT using $\beta_p$ and $n_p$. The lower BIC for the expanded model indicates that the improved fit is sufficient to overcome the penalty for having two extra parameters. This supports the idea that RT is still largely contaminated by movement processing because the model which directly deals with MPTs has the lower BIC.

A fourth comparison is between FProgReacR and CProgReacR. The latter constrains the $n_p$ parameter for MPT in the expanded model to match $n$ from the best model of MT. The reduction in BIC for the constrained model comes from decreased flexibility in how the parameters adapt to the data. The $\chi$ and $\chi^2$ values are nearly the same for the two models; the only difference between the model fittings is that $n_p$ is fixed in CProgReacR after being optimized for a model of MT. The two step fitting creates a hidden complexity for which the BIC does not penalize the model. Although most of the parameter values of the two models are similar, finding the best value for the maximum number of submovements for MT eliminates computation time in finding a good value of the parameter for MPT later.

The final comparison is between CProgReacR and CProgReacT. The same expanded model was fitted to rectangular and triangular conditions separately. All the fit statistics improve for the fit to triangular conditions, but the $\chi$ values indicate that the improvement may be minuscule when the amount of data is taken into account. Although the improvement in fitting the basic model from rectangular to triangular conditions was notable, the same may not be true for fitting the expanded model.

In the final chapter, I draw some general conclusions about the modeling of dynamic confidence and discuss the limitations of the study.
Chapter 5: GENERAL DISCUSSION

The purpose of this thesis was to explain how the four variables of choice, confidence, response time, and movement time relate to one another in a decision task. The main focus was on how the manner in which confidence ratings are collected impacts response times. Movement and preparation for movement contribute to response times, but their contribution may not be the same under all circumstances. The project was devoted to studying how task characteristics that affect movement time (MT) also affect response time. Primarily, I was attentive to the method of eliciting confidence ratings in experiments, which is often neglected in studying confidence response times.

To study the effects of the instruments of data collection, I set up multiple versions of a decision task. The experiment had two factors: one that dictated whether or not participants gave confidence ratings and one that changed the display of touch screen areas from rectangular to triangular. When people did not provide confidence ratings, they performed a simple choice task in which responses were to targets specified by the experimenter. In those targeting conditions, people provided reaction times (RTs) for choices and MTs to targets, so that without confidence judgments, decisions were faster and any effects of targets on MTs were isolated from effects on RTs.
In the rectangular display conditions of the experiment, participants used a response input display that Fitts’s Law and the optimized-submovement model (OSM) predicted would result in different MTs for different target/confidence levels. In the triangular display conditions, participants used a response input display that the models predicted would result in similar MTs for all target/confidence levels.

The full model I’ve proposed was able to account for the overall increase in response time from targeting to confidence conditions by allowing the decision threshold to vary (as suggested by Usher et al., 2002). To account for increases in MT as target distance increased, the model incorporated the OSM. The same model structure was used to account for movement programming times (MPTs) as well.

Furthermore, I showed that with careful design, response times can be insulated from the impact of MTs. In the confidence conditions, people provided response times and MTs which depended partly on confidence. When participants provided confidence ratings on the rectangular touchscreen areas, the influence of the display and confidence on response time were intentionally confounded as a demonstration of typical experimental methods. Response time decreased as confidence increased, but response time also increased as MT increased. In the targeting conditions, it was clear that rectangular touchscreen areas and triangular touchscreen areas had different effects on MT. Some participants showed no change in MTs with changes in target distance in the triangular condition. That is, the triangular areas released response time from the influence of movement in targeting conditions.

This research has both a strong methods component and a strong modeling component. The remainder of the discussion is devoted to the success of the experiment.
and the full model, followed by the limitations of both. Finally I summarize how the objectives of the research have been met.

5.1 Experiment Results

What most decision-making researchers would like to measure are decision times (DTs), but they must settle for response times. Often they assume that within the measurements of response time is a constant extra time. The primary motivation of this study was to evaluate whether that assumption was reasonable. Movement is a large part of the extra time included in total response time that may also interact with RTs, from which MTs have been subtracted. In decision tasks, the potential influence of movement on RT lies in MPT, which cannot be cleanly removed from RTs. Decision making and movement programming must occur prior to movement while no overt behavior shows them. My full model assumed that MPTs are impacted by the same factors as MTs, so that Fitts’ Law or the OSM can be applied to both MPTs and MTs. By this logic, constructing an input display for confidence that capitalizes on Fitts’ Law should make all MTs and MPTs roughly equal so that DTs can be modeled more directly.

The results of this approach were mixed. Overall, MTs were more similar across targets and confidence ratings for the triangular conditions of the experiment than for the rectangular conditions of the experiment. While using the rectangular input display, all participants’ MTs increased with target distance and confidence level as predicted by the OSM. However, the triangular input display eliminated the increase in MTs. In the triangular conditions, the participants showed little variation in MTs for targeting but showed some decreases in MTs for increases in confidence rating. For
the latter, recalculating the indexes of difficulty for movement with effective target widths and distances rather than nominal target widths and distances did not explain the results.

RTs had a correlation with MTs in the rectangular conditions, but the correlation was reduced in the triangular conditions. The nonsignificance of the latter correlation suggests that changes to the response input display can negate the impact of MPTs within RTs. Researchers concerned about movement-process contamination of response times could use my method to avoid the problem.

5.2 Model Evaluation

With its extra free parameter, \( n \), the OSM accounted for the MTs in the experiment better than Fitts’ Law. The values of the BIC confirmed that the added complexity of the model was worthwhile. Many of the unusual patterns of MTs could be accommodated by either model by allowing their parameters to change for different people in different conditions, which suggests that ways MT deviated from Fitts’ Law or the OSM on the surface were rooted in peculiar behaviors of the experiment participants between sessions.

The ability to alter the influence of the input display on response time was confirmed by the improved fit of the basic model of total response time in triangular conditions (BasicRespT). This model did not incorporate the OSM, yet for triangular conditions it fitted the data better than in other conditions. The same was true for the fit of the basic model to RT: the basic model fitted RTs better than response times in the rectangular conditions. These outcomes mean that subtracting MT from response time improved the fit of the basic model, and altering the computer interface
for participants improved the fit of the basic model. The improvement suggests that where the experimental results for response time in the triangular confidence condition were qualitatively different from results for response time in the rectangular confidence condition, the model was able to take advantage of differences.

In more typical experimental methods for confidence, no consideration is given to MTs, so I expected that the typical input display for confidence also confounds MPTs with RTs. The model evaluation showed that dynamic choice and confidence models should not be applied to response times without first separating MPTs from RTs. This requires increasing the number of free parameters in the model, but the values of the model evaluation criteria indicate that it is worth the increase in model complexity. All of the expanded models outperformed their basic model counterparts by incorporating the OSM for MPTs. Model developers who assume a baseline for response time by attaching a constant non-decision component ignore the opportunity for improvement in fits to the decision process.

Complexity was a concern for the model evaluation because the expanded models had a different form that included extra free parameters. The BIC was used as the model evaluation criterion because it penalizes models for their number of free parameters. Despite increased complexity, the expanded models outperformed the basic models. More parameters were need to model MPT, but they improved models of RT sufficiently to justify their use. The BIC was lower for the expanded model in a direct comparison with the basic model (FProgReacR and BasicReacR fittings).
5.3 Limitations

It appears that small targets produce more individual differences in MTs and MPTs, so future studies could alter the input display for confidence to compensate more effectively for behavior involving MTs and MPTs. If the smaller targets at the shorter distances were removed, participants might perform the targeting task more uniformly. Confidence would have to be rescaled for fewer discrete ratings (perhaps five instead of seven), which would encourage participants to use the full new scale.

An added benefit to decreasing the number of confidence levels would be allowing more space for the rectangular targets to expand. Their moderate size in the foregoing experiment made the index of difficulty in the rectangular conditions consistently higher than in the triangular conditions. A more direct comparison between MTs would be possible if the index of difficulty for the triangular conditions were in the middle of the range of index of difficulty for the rectangular conditions. Increasing the size of the targets in the rectangular conditions would allow this comparison.

The reduction in number of unique confidence ratings would have yet another benefit of reducing the size of the array used for model fitting, which in turn would reduce the fit statistics and speed up parameter optimization.

A major limitation of the modeling is the amount of data being described. 200 trials per person per condition seems sufficient until one notices that participants provided very few of the lower confidence ratings. In optimizing the models, up to 800 data points were placed in 140 bins, which means a bin typically had six predicted data in it, but could have had as few as one. Although this rarely occurred, it presents a problem for model fitting because $\chi^2$ may not be robust to small cell counts. Even if each bin had more than five data points, the reliability of the parameter estimates
could suffer. A different experimental design might be able to gather more data per person without overtaxing participants.

The $\chi^2$ was chosen as the statistic to minimize during parameter optimization, but others could be used. However, any measures of fit that require a likelihood would be impossible when using simulations of the model. In the future, the model’s full joint probability density function could be developed, and maximum likelihood estimation could be implemented for model selection. Computation would be intense, but behavior of the model would be smoother over changes in parameter values.

The BIC was used with an approximate likelihood for model evaluation. The rule for comparing models’ BICs is that the model with the lower value of the BIC is preferred. When models are fitted to the same data, such as the fits of the basic and expanded models to the RTs in the rectangular condition, a direct comparison between models is possible. In this case, the expanded model was better, but only by a difference of 56, less than one percent of the BIC for the basic model. Even if about one quarter of the values of BICs are discounted because they are shared due to the essential number of parameters and data, the improvement in model fit is relatively small. This problem with the BIC is not unique. Many model evaluation criteria don’t have formalized methods for concluding how much one model is better than another — only that it is.

Although accuracy of choices is important in decision research, it was not the primary focus of this project. Thus, it was not carefully examined during model fitting. Future modeling should consider accuracy more fully as it has long been a concern to testers of dynamic choice models.
Future model testing could start with accuracy of choices. It is important to obtain realistic RT distributions and confidence distributions for correct and error responses. In the present study, I was not looking to falsify the models but rather to show relative quality of explanation between the models for different results. The models could be contested with data that would also discredit other dynamic confidence models, but a search of the model predictions over the whole parameter space might find other types of data that could uniquely falsify the models I’ve created for this research. The newest characteristic of the models is an account of MPT within choice RT. Data in which MPTs are not proportional to MTs would falsify the expanded model of choice/confidence RT, and this data should be sought in the future.

The data in this experiment indicate that the dependency between MT and RT may run both directions, something the model does not account for explicitly. By allowing the parameters to vary across conditions, the OSM and Fitts’ Law were able to increase MTs for confidence conditions, in which RTs were greater than in targeting conditions. These parameter changes do not provide an explanation for how increasing RT increases MT. Instead, the expanded model is based on the theory that increasing MT increases RT by a small but significant amount.

5.4 Conclusion

I set out to discover whether parameters in choice response time models for constant movement times were realistic. To test the idea, I had to meet three main objectives. First was to gather data on choices, confidence ratings, response times, and movement times for developing a dynamic confidence model. The BAC was created for explaining those variables as measured in the experiment.
The second objective was to isolate the influence of MT on the model. This was accomplished by evaluating a set of related models in which some more elaborately accounted for MPT within RT than others. The expanded models did perform better than the basic models. Not only does this suggest that RT is not independent of MT in many experimental paradigms, it suggests that MT and MPT are worth considering when modeling data from what is primarily a decision task.

The value of explicitly accounting for motor responses has major implications for confidence and choice response time research. When researchers cite a difference in choice response time of 30 ms, it could be an artifact of the way participants move to make responses. Also, when a dynamic choice model fits a response time distribution better than another model, it could be fitting components of the task other than decision processing. The shape of the distribution could be influenced by movement in subtle ways. By not having the model structure to account for MTs or MPTs, the model’s other parameters may be incorrectly estimated. Because parameters generally carry psychological meaning, parameter estimates could be misinterpreted.

The third objective was to make consideration of MT and MPT in the model less important when they are instead considered in the experimental methods. I was able to at least partially compensate for predictable variation in MT and its influence on RT by constructing the apparatus and input display for confidence according to Fitts’s Law. The basic model fitted data better for triangular conditions than for the more traditional rectangular conditions even without the model structure to account for variability in MTs or MPTs.
Appendix A: APPENDIX

A.1 Formulas

The CDF for $z$ at time $t$ begins with the exponential PDF for an accumulation rate, $d_i$:

$$
\begin{cases}
v_i e^{-v_i d_i} & \text{if } d_i \geq 0 \\
0 & \text{otherwise.}
\end{cases}
$$

From this, I obtain the CDF for an accumulation rate:

$$
\begin{cases}
1 - e^{-v_i d_i} & \text{if } d_i \geq 0 \\
0 & \text{otherwise.}
\end{cases}
$$

I also need the uniform PDF for $a$, the starting point of the accumulators:

$$
\begin{cases}
\frac{1}{A} & \text{if } 0 \leq a \leq A \\
0 & \text{otherwise.}
\end{cases}
$$

The CDF for evidence on an accumulator is the probability that the process has traveled less than $z$ distance, which is dependent on the rate of travel and the starting point, both of which are random.

$$p(a + d_i t < z) = p \left( d_i < \frac{z - a}{t} \right) p(a)$$

The first term can be rewritten using the CDF of $d_i$: $1 - e^{-v_i (z-a)/t}$. The second term can be rewritten using the PDF of $a$: $1/A$. After combining them, I integrate over
the possible values of $a$; however, because the first term contains $(z - a)/t$ in place of $d_i$, $z$ must be greater than or equal to $A$ to integrate up to $A$.

$$p(a + d_i t < z) = \begin{cases} 
\int_0^A (1 - e^{v_i(z-a)/t}) \frac{1}{A} da & \text{if } z \geq A \\
\int_0^z (1 - e^{v_i(z-a)/t}) \frac{1}{A} da & \text{if } 0 < z < A.
\end{cases}$$  \hspace{1cm} (A.1)$$

To avoid replacing $d_i$ with a negative value, I integrate only from 0 to $z$ when $z$ is less than $A$.

I need to find the indefinite integral for Eq. A.1. Starting with a simpler case where $K$ is a constant,

$$\int 1 - e^{-Ka} da = a - e^{-Ka} \left( -\frac{1}{K} \right) = a + \frac{e^{-Ka}}{K}.$$  

For the more complicated case,

$$\int (1 - e^{-v_i(z-a)/t}) \frac{1}{A} da = \int (1 - e^{(-v_i z/t + v_i a/t)}) \frac{1}{A} da = \left( a - \frac{t}{v_i} e^{-v_i(z-a)/t} \right) \frac{1}{A}.$$  

After rearranging terms, the indefinite integral becomes

$$v_i a - te^{-v_i(z-a)/t}$$

$$v_i A.$$

Using the first fundamental theorem of calculus, I plug in the upper and lower limits for $a$ and find the difference.

$$\int_0^A (1 - e^{-v_i(z-a)/t}) \frac{1}{A} da = \frac{v_i A - te^{-v_i(z-A)/t}}{v_i A} - \frac{v_i \times 0 - te^{-v_i(z-0)/t}}{v_i A}$$

$$= 1 - \frac{t \left( e^{-v_i(z-A)/t} - e^{-v_i(z-0)/t} \right)}{v_i A}$$

$$= 1 + te^{-v_i z/t} \left( 1 - e^{v_i A/t} \right) \frac{1}{v_i A}.$$  

$$\int_0^z (1 - e^{-v_i(z-a)/t}) \frac{1}{A} da = \frac{v_i z - te^{-v_i(z-z)/t}}{v_i A} - \frac{v_i \times 0 - te^{-v_i(z-0)/t}}{v_i A}$$

$$= \frac{v_i z - t \left( e^{-v_i z/t} \right)}{v_i A}$$

$$= \left( -t \left( 1 - e^{-v_i z/t} \right) + v_i z \right) \frac{1}{v_i A}.$$
To find the PDF for \( z \) at time \( t \), the CDF must be differentiated with respect to \( z \). Again, starting with a simpler case where \( K \) is a constant,

\[
\frac{d}{dz} e^{-Kz} = -Ke^{-Kz}.
\]

Hence,

\[
\frac{d}{dz} (1 - e^{-v_i z/t}) = -\frac{v_i}{t} e^{-v_i z/t},
\]

and

\[
\frac{d}{dz} \left( -t (1 - e^{-v_i z/t}) + v_i z \right)/(v_i A) = (1 - e^{-v_i z/t}) \frac{1}{A}.
\]

This result should be obvious from the second piece of Eq. A.1. The result I desire for the first piece is less apparent. This part of the CDF can be rewritten as

\[
1 + \frac{t \left( e^{-v_i z/t} - e^{-v_i(z-A)/t} \right) \frac{1}{v_i A}}{1 - e^{-v_i z/t} - e^{-v_i(z-A)/t}}.
\]

Because

\[
\frac{d}{dz} \left( t (e^{-v_i z/t} - e^{-v_i(z-A)/t}) \right) = -v_i e^{-v_i z/t} + v_i e^{-v_i(z-A)/t},
\]

it is also true that

\[
\frac{d}{dz} \left( 1 + \frac{t \left( e^{-v_i z/t} - e^{-v_i(z-A)/t} \right) \frac{1}{v_i A}}{1 - e^{-v_i z/t} - e^{-v_i(z-A)/t}} \right) = (e^{-v_i(z-A)/t} - e^{-v_i z/t}) \frac{1}{A}
\]

\[
= -e^{-v_i z/t} (1 - e^{v_i A/t}) \frac{1}{A}.
\]

Putting the two results together provides Eq. 2.4:

\[
s_i(z) = \begin{cases} 
- e^{-v_i z/t} (1 - e^{v_i A/t}) \frac{1}{A} & \text{if } z \geq A \\
(1 - e^{-v_i z/t}) \frac{1}{A} & \text{if } 0 < z < A. 
\end{cases}
\]

As stated in the second chapter, the CDF for the time when an accumulator reaches threshold \( b \) is one minus the CDF for evidence \( z \) when \( z = b \). Because \( b \) is assumed to be greater than \( A \), the solution is simple. As Eq. 2.5 showed,

\[
G_i(t) = -te^{-v_i b/t} (1 - e^{v_i A/t}) \frac{1}{v_i A}.
\]
Through the same process as before, I can differentiate the function, but now with respect to $t$:

$$
\frac{dG_i}{dt} = \left( (v_i b + t)(e^{-v_i (b-A)/t} - e^{-v_i b /t}) - v_i Ae^{-v_i (b-A)/t} \right) \frac{1}{v_i At}.
$$

This gives the PDF for decision times as seen in Eq. 2.6.
A.2 Instructions to Participants for Targeting Conditions

In this experiment, you will try to discriminate between different climates of a country, one city at a time. When you press and hold the plus sign, a number will appear in the middle of the screen. The number tells the inches of rain measured last year for a city in one of four climate regions, and you need to decide which region the city belongs to. Region D is a Dry grassland with Very low rainfall; region S is a Savanna with low rainfall; region T is a Tropical forest with high rainfall; and region R is a Rain forest with Very high rainfall.

In this first practice phase, you will learn how to pick a climate region: numbers from region D have an average of 20; numbers from region S have an average of 40; numbers from region T have an average of 60; numbers from region R have an average of 80. To choose the climate region you think made the number in the middle of the screen, release the plus sign and use the same finger to touch the white bar with that letter. As soon as you make a choice, the computer will say which was the correct region. Keep in mind that each year a city can have different rainfall from average, so occasionally a tricky number will appear. For example, a low amount of rain could be from region T.

30 training trials

In this second practice phase, you will have an added task involving response targets. The choice bars are arranged like spokes on a wheel. Before a number shows up, your finger is in the center of the bars because you don’t yet know which choice to make. Furthermore, you don’t know how far to move your finger. The bars are now divided into levels that may be small distances or large distances from the center. The levels will all be shown together for now, but you should touch just one level
on the bar of your choice. Try touching a different level for each choice to become accustomed to targeting both low levels and high levels.

30 training trials

Only release the plus sign after you know the point where you want to touch. The practice you just completed showed how all the levels of the bars fit together and let you pick which levels to touch. The lower levels feel different from the higher levels. In the third practice phase you will be asked to touch specific levels when you make your choices. On average, you will see a medium target level from a mix of low and high target levels. To help you track how well you’re doing, you will start earning money for your answers. Correct choices earn positive cents, and errors earn negative cents. The number of cents gained or lost will be based on targeted levels. Low levels can add or subtract low numbers of cents, and high levels can add or subtract high numbers of cents. To earn the most money, you should make correct choices and accurate finger movements.

30 training trials

Now that practice is over, there will be four large sets of cities. You will continue to see how many cents you earn for each choice. Please try to make your choices and finger movements accurately and quickly. Remember to not lift your finger off the plus sign until you know where you want to press down again. Ask the experimenter now if you have any questions.

A.3 Instructions to Participants for Confidence Conditions

In this experiment, you will try to discriminate between different climates of a country, one city at a time. When you press and hold the plus sign, a number will
appear in the middle of the screen. The number tells the inches of rain measured last year for a city in one of four climate regions, and you need to decide which region the city belongs to. Region D is a Dry grassland with Very low rainfall; region S is a Savanna with low rainfall; region T is a Tropical forest with high rainfall; and region R is a Rain forest with Very high rainfall.

In this first practice phase, you will learn how to pick a climate region: numbers from region D have an average of 20; numbers from region S have an average of 40; numbers from region T have an average of 60; numbers from region R have an average of 80. To choose the climate region you think made the number in the middle of the screen, release the plus sign and use the same finger to touch the white bar with that letter. As soon as you make a choice, the computer will say which was the correct region. Keep in mind that each year a city can have different rainfall from average, so occasionally a tricky number will appear. For example, a low amount of rain could be from region T.

30 training trials

In this second practice phase, you will have an added task involving confidence. The choice bars are arranged like spokes on a wheel. Before a number shows up, your finger is in the center of the bars because you don’t yet know which choice to make. The center is for complete uncertainty; the outer edges are for complete certainty. Touching farther from the center means you are more certain about which climate region the city is in. Imagine a city has 74 inches of rain. It probably is in region R, but there’s a chance that it came from region T. You feel mostly confident about choosing the Rain forest, so you should touch a point about two-thirds of the distance
between the center and letter R. Only release the plus sign after you know the point where you want to touch.

30 training trials

The practice you just completed included many numbers that were easy to place into climate regions. When you were making the right choices, you should have pressed on the outer ends of the choice bars to show that you were confident. In the third practice phase you will see a mix of hard and easy rainfall numbers, so you should have medium confidence on average. To help you track how well you’re doing, you will start earning money for your answers. Correct choices earn positive cents, and errors earn negative cents. The number of cents gained or lost will be based on confidence. Low confidence can add or subtract low numbers of cents, and high confidence can add or subtract high numbers of cents. To earn the most money, you should press for high confidence on correct choices and press for low confidence on incorrect choices.

30 training trials

Now that practice is over, there will be four large sets of cities. You will continue to see how many cents you earn for each choice. Please try to make your choices in line with your confidence accurately and quickly. Remember to not lift your finger off the plus sign until you know where you want to press down again. Ask the experimenter now if you have any questions.

A.4 Model Fits
<table>
<thead>
<tr>
<th>Model &amp; Data</th>
<th>Estimate</th>
<th>Participant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>BasicRespR</td>
<td>range start</td>
<td>A</td>
</tr>
<tr>
<td>Without program</td>
<td>threshold</td>
<td>b</td>
</tr>
<tr>
<td>Response time</td>
<td>criterion</td>
<td>c_4</td>
</tr>
<tr>
<td>Rectangular</td>
<td>criterion</td>
<td>c_5</td>
</tr>
<tr>
<td></td>
<td>criterion</td>
<td>c_6</td>
</tr>
<tr>
<td></td>
<td>criterion</td>
<td>c_7</td>
</tr>
<tr>
<td></td>
<td>criterion</td>
<td>c_8</td>
</tr>
<tr>
<td></td>
<td>criterion</td>
<td>c_9</td>
</tr>
<tr>
<td></td>
<td>min NDT</td>
<td>α</td>
</tr>
<tr>
<td></td>
<td>goodness of fit</td>
<td>χ²</td>
</tr>
<tr>
<td></td>
<td>evaluation</td>
<td>BIC</td>
</tr>
<tr>
<td>BasicRespT</td>
<td>range start</td>
<td>A</td>
</tr>
<tr>
<td>Without program</td>
<td>threshold</td>
<td>b</td>
</tr>
<tr>
<td>Reponse time</td>
<td>criterion</td>
<td>c_4</td>
</tr>
<tr>
<td>Triangular</td>
<td>criterion</td>
<td>c_5</td>
</tr>
<tr>
<td></td>
<td>criterion</td>
<td>c_6</td>
</tr>
<tr>
<td></td>
<td>criterion</td>
<td>c_7</td>
</tr>
<tr>
<td></td>
<td>criterion</td>
<td>c_8</td>
</tr>
<tr>
<td></td>
<td>criterion</td>
<td>c_9</td>
</tr>
<tr>
<td></td>
<td>min NDT</td>
<td>α</td>
</tr>
<tr>
<td></td>
<td>goodness of fit</td>
<td>χ²</td>
</tr>
<tr>
<td></td>
<td>evaluation</td>
<td>BIC</td>
</tr>
<tr>
<td>BasicReacR</td>
<td>range start</td>
<td>A</td>
</tr>
<tr>
<td>Without program</td>
<td>threshold</td>
<td>b</td>
</tr>
<tr>
<td>Reaction time</td>
<td>criterion</td>
<td>c_4</td>
</tr>
<tr>
<td>Rectangular</td>
<td>criterion</td>
<td>c_5</td>
</tr>
<tr>
<td></td>
<td>criterion</td>
<td>c_6</td>
</tr>
<tr>
<td></td>
<td>criterion</td>
<td>c_7</td>
</tr>
<tr>
<td></td>
<td>criterion</td>
<td>c_8</td>
</tr>
<tr>
<td></td>
<td>criterion</td>
<td>c_9</td>
</tr>
<tr>
<td></td>
<td>min MPT</td>
<td>α_p</td>
</tr>
<tr>
<td></td>
<td>goodness of fit</td>
<td>χ²</td>
</tr>
<tr>
<td></td>
<td>evaluation</td>
<td>BIC</td>
</tr>
</tbody>
</table>

Table continues on next page.
|
|---------------------------------|------------------|------------------|------------------|------------------|------------------|
| **Model & Data**                | **Estimate**     | **Participant**  |
| FProgReaR range start           | A 3.47 4.09 6.79 | 1 2 3 4 Group    |
| $n_p$ is free threshold         | b 9.27 13.1 10.2 |                 |
| Reaction time criterion         | $c_4$ 0.379 0.351 |                 |
| Rectangular criterion           | $c_5$ 0.408 0.398 |                 |
| criterion                       | $c_6$ 0.439 0.448 |                 |
| criterion                       | $c_7$ 0.487 0.507 |                 |
| criterion                       | $c_8$ 0.508 0.590 |                 |
| criterion                       | $c_9$ 0.554 0.679 |                 |
| min MPT $\alpha_p$             | 53.4 100 47.9 51.1|                 |
| MPT slope $\beta_p$            | 47.3 33.2 23.1 38.0|                 |
| submovements $n_p$             | 4.22 5.81 3.43 4.28|                 |
| goodness of fit $\chi^2$       | 317 254 679 306 261|                 |
| evaluation BIC                 | 12160            |                 |
| CProgReaR range start           | A 3.40 4.16 6.15 | 1 2 3 4 Group    |
| $n_p$ from OSM threshold        | b 7.74 10.17 11.85|                 |
| Reaction time criterion         | $c_4$ 0.412 0.441 |                 |
| criterion                       | $c_5$ 0.443 0.494 |                 |
| criterion                       | $c_6$ 0.475 0.545 |                 |
| criterion                       | $c_7$ 0.529 0.607 |                 |
| criterion                       | $c_8$ 0.551 0.687 |                 |
| criterion                       | $c_9$ 0.602 0.776 |                 |
| min MPT $\alpha_p$             | 16.3 31.3 22.1 16.9|                 |
| MPT slope $\beta_p$            | 65.7 106 96.8 100|                 |
| submovements $n_p$             | 3.74 5.67 2.98 3.60|                 |
| goodness of fit $\chi^2$       | 269 261 688 527 259|                 |
| evaluation BIC                 | 12124            |                 |
| CProgReaCT range start          | A 4.74 2.51 5.36 | 1 2 3 4 Group    |
| $n_p$ from OSM threshold        | b 9.42 8.48 10.77|                 |
| Reaction time criterion         | $c_4$ 0.231 0.318 |                 |
| criterion                       | $c_5$ 0.242 0.373 |                 |
| criterion                       | $c_6$ 0.384 0.416 |                 |
| criterion                       | $c_7$ 0.425 0.465 |                 |
| criterion                       | $c_8$ 0.486 0.512 |                 |
| criterion                       | $c_9$ 0.519 0.599 |                 |
| min MPT $\alpha_p$             | 29.0 23.3 7.34 35.7|                 |
| MPT slope $\beta_p$            | 34.5 32.1 46.7 30.8|                 |
| goodness of fit $\chi^2$       | 235 167 649 206 236|                 |
| evaluation BIC                 | 11269            |                 |

Table A.1: Parameter estimates from model fits. The table shows parameter estimates averaged across conditions and fit statistics for six dynamic confidence models.
<table>
<thead>
<tr>
<th>Model</th>
<th>Condition</th>
<th>Estimate</th>
<th>Participant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( \alpha_m )</td>
<td>1</td>
</tr>
<tr>
<td>Fitts</td>
<td>Rect Targeting</td>
<td>min MT</td>
<td>3.11</td>
</tr>
<tr>
<td>Nominal</td>
<td>MT slope</td>
<td>( \beta_m )</td>
<td>57.7</td>
</tr>
<tr>
<td></td>
<td>standard dev</td>
<td>( s_m )</td>
<td>47.2</td>
</tr>
<tr>
<td>Tria Targeting</td>
<td>min MT</td>
<td>( \alpha_m )</td>
<td>163</td>
</tr>
<tr>
<td></td>
<td>MT slope</td>
<td>( \beta_m )</td>
<td>7.27</td>
</tr>
<tr>
<td></td>
<td>standard dev</td>
<td>( s_m )</td>
<td>49.3</td>
</tr>
<tr>
<td>Rect Confidence</td>
<td>min MT</td>
<td>( \alpha_m )</td>
<td>43.3</td>
</tr>
<tr>
<td></td>
<td>MT slope</td>
<td>( \beta_m )</td>
<td>46.5</td>
</tr>
<tr>
<td></td>
<td>standard dev</td>
<td>( s_m )</td>
<td>46.5</td>
</tr>
<tr>
<td>Tria Confidence</td>
<td>min MT</td>
<td>( \alpha_m )</td>
<td>155</td>
</tr>
<tr>
<td></td>
<td>MT slope</td>
<td>( \beta_m )</td>
<td>11.5</td>
</tr>
<tr>
<td></td>
<td>standard dev</td>
<td>( s_m )</td>
<td>51.3</td>
</tr>
<tr>
<td></td>
<td>goodness of fit</td>
<td>( \chi^2 )</td>
<td>220</td>
</tr>
<tr>
<td></td>
<td>evaluation BIC</td>
<td></td>
<td>2598</td>
</tr>
<tr>
<td>Fitts</td>
<td>Rect Targeting</td>
<td>min MT</td>
<td>54.5</td>
</tr>
<tr>
<td>Effective</td>
<td>MT slope</td>
<td>( \beta_m )</td>
<td>59.1</td>
</tr>
<tr>
<td></td>
<td>standard dev</td>
<td>( s_m )</td>
<td>29.0</td>
</tr>
<tr>
<td>Tria Targeting</td>
<td>min MT</td>
<td>( \alpha_m )</td>
<td>167</td>
</tr>
<tr>
<td></td>
<td>MT slope</td>
<td>( \beta_m )</td>
<td>15.3</td>
</tr>
<tr>
<td></td>
<td>standard dev</td>
<td>( s_m )</td>
<td>53.3</td>
</tr>
<tr>
<td>Rect Confidence</td>
<td>min MT</td>
<td>( \alpha_m )</td>
<td>149</td>
</tr>
<tr>
<td></td>
<td>MT slope</td>
<td>( \beta_m )</td>
<td>43.1</td>
</tr>
<tr>
<td></td>
<td>standard dev</td>
<td>( s_m )</td>
<td>38.9</td>
</tr>
<tr>
<td>Tria Confidence</td>
<td>min MT</td>
<td>( \alpha_m )</td>
<td>169</td>
</tr>
<tr>
<td></td>
<td>MT slope</td>
<td>( \beta_m )</td>
<td>10.9</td>
</tr>
<tr>
<td></td>
<td>standard dev</td>
<td>( s_m )</td>
<td>49.7</td>
</tr>
<tr>
<td></td>
<td>goodness of fit</td>
<td>( \chi^2 )</td>
<td>371</td>
</tr>
<tr>
<td></td>
<td>evaluation BIC</td>
<td></td>
<td>25915</td>
</tr>
</tbody>
</table>

Table continues on next page.
<table>
<thead>
<tr>
<th>Model</th>
<th>Condition</th>
<th>Estimate</th>
<th>Participant</th>
<th></th>
<th></th>
<th></th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSM</td>
<td>Rect Targeting</td>
<td>min MT</td>
<td>$\alpha_m$</td>
<td>7.31</td>
<td>92.4</td>
<td>65.3</td>
<td>17.4</td>
</tr>
<tr>
<td>Nominal</td>
<td>MT slope</td>
<td>$\beta_m$</td>
<td>181</td>
<td>238</td>
<td>192</td>
<td>210</td>
<td>205</td>
</tr>
<tr>
<td></td>
<td>submovements</td>
<td>$n_m$</td>
<td>4.02</td>
<td>5.90</td>
<td>3.11</td>
<td>3.79</td>
<td>4.22</td>
</tr>
<tr>
<td></td>
<td>standard dev</td>
<td>$s_m$</td>
<td>39.6</td>
<td>53.1</td>
<td>49.5</td>
<td>58.1</td>
<td>50.1</td>
</tr>
<tr>
<td></td>
<td>Tria Targeting</td>
<td>min MT</td>
<td>$\alpha_m$</td>
<td>155</td>
<td>253</td>
<td>223</td>
<td>294</td>
</tr>
<tr>
<td></td>
<td>MT slope</td>
<td>$\beta_m$</td>
<td>20.5</td>
<td>23.1</td>
<td>34.0</td>
<td>15.6</td>
<td>237</td>
</tr>
<tr>
<td></td>
<td>submovements</td>
<td>$n_m$</td>
<td>2.06</td>
<td>2.15</td>
<td>1.84</td>
<td>2.51</td>
<td>2.14</td>
</tr>
<tr>
<td></td>
<td>standard dev</td>
<td>$s_m$</td>
<td>36.8</td>
<td>67.5</td>
<td>61.2</td>
<td>48.5</td>
<td>53.5</td>
</tr>
<tr>
<td>Rect Confidence</td>
<td>min MT</td>
<td>$\alpha_m$</td>
<td>51.3</td>
<td>90.1</td>
<td>57.3</td>
<td>81.6</td>
<td>50.1</td>
</tr>
<tr>
<td></td>
<td>MT slope</td>
<td>$\beta_m$</td>
<td>164</td>
<td>279</td>
<td>242</td>
<td>261</td>
<td>23.3</td>
</tr>
<tr>
<td></td>
<td>submovements</td>
<td>$n_m$</td>
<td>3.66</td>
<td>5.39</td>
<td>3.15</td>
<td>3.61</td>
<td>3.95</td>
</tr>
<tr>
<td></td>
<td>standard dev</td>
<td>$s_m$</td>
<td>38.1</td>
<td>50.9</td>
<td>54.3</td>
<td>43.9</td>
<td>46.8</td>
</tr>
<tr>
<td>Tria Confidence</td>
<td>min MT</td>
<td>$\alpha_m$</td>
<td>149</td>
<td>251</td>
<td>241</td>
<td>323</td>
<td>241</td>
</tr>
<tr>
<td></td>
<td>MT slope</td>
<td>$\beta_m$</td>
<td>30.8</td>
<td>23.1</td>
<td>42.3</td>
<td>18.5</td>
<td>28.7</td>
</tr>
<tr>
<td></td>
<td>submovements</td>
<td>$n_m$</td>
<td>2.28</td>
<td>1.95</td>
<td>1.75</td>
<td>2.21</td>
<td>2.05</td>
</tr>
<tr>
<td></td>
<td>standard dev</td>
<td>$s_m$</td>
<td>40.5</td>
<td>59.9</td>
<td>60.2</td>
<td>58.7</td>
<td>54.8</td>
</tr>
<tr>
<td></td>
<td>goodness of fit</td>
<td>$\chi^2$</td>
<td>201</td>
<td>179</td>
<td>183</td>
<td>192</td>
<td>274</td>
</tr>
<tr>
<td>OSM</td>
<td>Rect Targeting</td>
<td>min MPT</td>
<td>$\alpha_m$</td>
<td>-23.9</td>
<td>0.88</td>
<td>33.3</td>
<td>-20.1</td>
</tr>
<tr>
<td>Effective</td>
<td>MT slope</td>
<td>$\beta_m$</td>
<td>158</td>
<td>258</td>
<td>167</td>
<td>155</td>
<td>184.5</td>
</tr>
<tr>
<td></td>
<td>submovements</td>
<td>$n_m$</td>
<td>3.89</td>
<td>6.88</td>
<td>1.69</td>
<td>3.23</td>
<td>3.92</td>
</tr>
<tr>
<td></td>
<td>standard dev</td>
<td>$s_m$</td>
<td>37.3</td>
<td>51.7</td>
<td>54.5</td>
<td>55.1</td>
<td>49.7</td>
</tr>
<tr>
<td>Tria Targeting</td>
<td>min PT</td>
<td>$\alpha_m$</td>
<td>156</td>
<td>254</td>
<td>231</td>
<td>291</td>
<td>233</td>
</tr>
<tr>
<td></td>
<td>MT slope</td>
<td>$\beta_m$</td>
<td>11.3</td>
<td>24.9</td>
<td>39.3</td>
<td>16.9</td>
<td>23.1</td>
</tr>
<tr>
<td></td>
<td>submovements</td>
<td>$n_m$</td>
<td>2.27</td>
<td>2.29</td>
<td>3.07</td>
<td>3.02</td>
<td>2.66</td>
</tr>
<tr>
<td></td>
<td>standard dev</td>
<td>$s_m$</td>
<td>34.0</td>
<td>50.5</td>
<td>51.5</td>
<td>48.3</td>
<td>46.1</td>
</tr>
<tr>
<td>Rect Confidence</td>
<td>min MT</td>
<td>$\alpha_m$</td>
<td>56.0</td>
<td>104</td>
<td>5.16</td>
<td>105</td>
<td>67.5</td>
</tr>
<tr>
<td></td>
<td>MT slope</td>
<td>$\beta_m$</td>
<td>101</td>
<td>189</td>
<td>169</td>
<td>119</td>
<td>144.5</td>
</tr>
<tr>
<td></td>
<td>submovements</td>
<td>$n_m$</td>
<td>3.45</td>
<td>5.33</td>
<td>1.60</td>
<td>2.76</td>
<td>3.29</td>
</tr>
<tr>
<td></td>
<td>standard dev</td>
<td>$s_m$</td>
<td>39.1</td>
<td>51.9</td>
<td>65.2</td>
<td>50.4</td>
<td>51.7</td>
</tr>
<tr>
<td>Tria Confidence</td>
<td>min MT</td>
<td>$\alpha_m$</td>
<td>158</td>
<td>250</td>
<td>243</td>
<td>321</td>
<td>243</td>
</tr>
<tr>
<td></td>
<td>MT slope</td>
<td>$\beta_m$</td>
<td>30.0</td>
<td>23.9</td>
<td>41.7</td>
<td>14.6</td>
<td>27.6</td>
</tr>
<tr>
<td></td>
<td>submovements</td>
<td>$n_m$</td>
<td>2.78</td>
<td>2.09</td>
<td>1.11</td>
<td>1.08</td>
<td>1.77</td>
</tr>
<tr>
<td></td>
<td>standard dev</td>
<td>$s_m$</td>
<td>43.7</td>
<td>66.3</td>
<td>52.6</td>
<td>55.8</td>
<td>54.6</td>
</tr>
<tr>
<td></td>
<td>goodness of fit</td>
<td>$\chi^2$</td>
<td>328</td>
<td>355</td>
<td>461</td>
<td>391</td>
<td>1187</td>
</tr>
<tr>
<td></td>
<td>evaluation</td>
<td>BIC</td>
<td>25345</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A.2: Parameter estimates from MT model fits. The table shows the parameter estimates and fit statistics for two versions of Fitts’s Law and the optimized-submovement model (OSM): nominal index of difficulty and effective index of difficulty.
References


Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.


