Perfectionism, Decision-Making, and Post-error Slowing

Dissertation

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

By

Kevin W. Potter, B.A., M.A.

Graduate Program in Psychology

The Ohio State University

2015

Dissertation Committee:

Trisha Van Zandt, Advisor

Amelia Aldao

Per Sederberg
Abstract

The construct of perfectionism is an important topic of research due to its influence on a large number of psychopathologies. Recently researchers have sought to better understand the cognitive mechanisms associated with perfectionism using forced-choice two-alternative tasks. In particular, perfectionist characteristics were expected to predict the behavioral measure of post-error slowing. However, research has found little association between inventory scores of perfectionism and post-error slowing. I developed a quantitative model for the Simon task, and tested several hypotheses regarding the mechanisms underlying post-error slowing and their association with perfectionism. I controlled for several confounding effects such as fatigue, motivation, learning, and multiple types of post-error adjustments. Critically, a core assumption past researchers have made regarding the interpretation of post-error slowing was found not to hold, leading to important implications regarding the study of the cognitive mechanisms underlying perfectionism.
This is dedicated to my family and loved ones.
Acknowledgments

I am indebted to Trish Van Zandt, my advisor, for her patience and guidance. I would also like to thank Susan, David, and Jill Potter, as well as Anne Weeks, for their help and suggestions while proofreading the early drafts of my thesis. I am grateful as well for the contributions by my fellow members in the Studies in Mathematical Psychology lab, Slava Nikitin and Bob Gore.
Vita

2009 ................................................. B.A. Psychology

2009-2010 ................................. University Fellow,  
The Ohio State University.

2010-2013 ................................. Graduate Teaching Associate,  
The Ohio State University.

2013-2014 ................................. Graduate Research Associate,  
The Ohio State University.

2014-Present ................................. Graduate Teaching Associate,  
The Ohio State University.

Publications

Research Publications


Fields of Study

Major Field: Psychology
# Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Perfectionism</td>
<td>3</td>
</tr>
<tr>
<td>1.1.1</td>
<td>Conceptualization of Perfectionism</td>
<td>3</td>
</tr>
<tr>
<td>1.1.2</td>
<td>Clinical Relevance</td>
<td>7</td>
</tr>
<tr>
<td>1.1.3</td>
<td>Neural Basis</td>
<td>12</td>
</tr>
<tr>
<td>1.2</td>
<td>Motivation and Personality</td>
<td>23</td>
</tr>
<tr>
<td>1.2.1</td>
<td>Motivation and the ERN.</td>
<td>23</td>
</tr>
<tr>
<td>1.2.2</td>
<td>Negative Emotionality and the ERN.</td>
<td>26</td>
</tr>
<tr>
<td>1.3</td>
<td>Post-error Behavior</td>
<td>29</td>
</tr>
<tr>
<td>1.3.1</td>
<td>Early Work on Post-error Slowing</td>
<td>30</td>
</tr>
<tr>
<td>1.3.2</td>
<td>Current Theories on Post-error Slowing</td>
<td>34</td>
</tr>
<tr>
<td>1.3.3</td>
<td>Support for Theories on Post-error Slowing</td>
<td>37</td>
</tr>
<tr>
<td>1.3.4</td>
<td>Risks with Aggregation</td>
<td>38</td>
</tr>
<tr>
<td>1.4</td>
<td>Quantitative Modeling</td>
<td>43</td>
</tr>
<tr>
<td>1.4.1</td>
<td>The Diffusion Race Model</td>
<td>47</td>
</tr>
<tr>
<td>1.5</td>
<td>Summary</td>
<td>52</td>
</tr>
</tbody>
</table>
5.4 Model Comparisons ........................................... 131
5.5 Posterior Predictive Checks ................................. 134
5.6 Associations with Perfectionism ............................ 144
  5.6.1 Group-level Effects .................................. 148
  5.6.2 Associations with “Concerns over Mistakes” .......... 150
  5.6.3 Associations with “Doubts about Actions” .......... 150
  5.6.4 Associations with the STAI Subscale ................. 152
  5.6.5 Correlations between Coefficients .................. 152
5.7 Further Analyses of Post-error Adjustments .............. 154
5.8 Summary ......................................................... 157

6. Discussion ......................................................... 161
  6.1 Implications .................................................. 163
    6.1.1 Post-Error Slowing ................................ 164
    6.1.2 Perfectionism and ERPs ............................ 165
    6.1.3 Adaptive Individual Deadlines .................... 167
    6.1.4 The Simon Task ................................... 169
  6.2 Methodological Extensions ............................... 176
    6.2.1 Hierarchical Models ............................... 176
    6.2.2 Estimation Routines ............................... 177
    6.2.3 Measurement of Perfectionism .................... 178
  6.3 Conclusions ................................................. 180

Appendices ..................................................................... 182
A. Model Fits for all Subjects .................................... 182

References ............................................................. 218
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>List of free parameters for all model variants.</td>
</tr>
<tr>
<td>4.1</td>
<td>A list of the lower and upper limits of the range used to generate random starting values from a uniform distribution, the boundaries between which each proposal fell, and the sampling block for each coefficient</td>
</tr>
<tr>
<td>5.1</td>
<td>Sessions and blocks removed based on percentage of timeouts, fast responses, and extremely poor performance.</td>
</tr>
<tr>
<td>5.2</td>
<td>The minimum and maximum Geweke diagnostic criterion values (z-scores) for each subject and model variant. More extreme values indicate large differences between the first 10% and the last 50% of the chain.</td>
</tr>
<tr>
<td>5.3</td>
<td>WAIC values for individual subject fits by model variant.</td>
</tr>
<tr>
<td>5.4</td>
<td>A summary of my hypotheses for the point estimates of the cognitive model and their group-level effects ($\beta_0$) as well as their associations with the “Concerns over mistakes” ($\beta_{CoM}$) and “Doubts about actions” ($\beta_{DaA}$) perfectionism subscales. Hypotheses labeled as confirmed had over 95% of the posterior mass above the predicted cutoff value. Hypotheses labeled as marginal had over 90% of the posterior mass over the cutoff point.</td>
</tr>
<tr>
<td>5.5</td>
<td>Correlations between point estimates for coefficients from the cognitive model, with variances on the diagonal. Asterisks indicate that the 95% credible intervals for the correlations do not overlap 0.</td>
</tr>
</tbody>
</table>
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>The accumulation process during a single trial for the diffusion race model given drift rates $\xi_i$ and $\xi_j$, thresholds $\kappa_i$ and $\kappa_j$ (set equal to each other for presentational clarity), and a non-decision component $\tau$ (a shift added to the decision time). Here, the accumulator for choice $i$, which has the higher drift rate, reaches its threshold first, producing the observed response time and subsequent choice of $i$.</td>
<td>49</td>
</tr>
<tr>
<td>2.1</td>
<td>The instructions, displayed stimuli, and timing using two example trials of the start of an experimental block.</td>
<td>57</td>
</tr>
<tr>
<td>3.1</td>
<td>The decision process in both the congruent (left-hand side) and incongruent (right-hand side) condition as proposed in my original model. The arrows denote the average path of the evidence accumulation, while the color of the arrow indicates to which choice the accumulator corresponds. The example assumes a red stimulus was shown.</td>
<td>65</td>
</tr>
<tr>
<td>3.2</td>
<td>Quantile-probability plot showing the 3rd pilot subjects’s observed 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles on the y-axis and probability of a correct or incorrect response on the x-axis for the congruent (circles) and incongruent (crosses) conditions. The 95% credible intervals for the model predictions are shown as ellipses.</td>
<td>67</td>
</tr>
<tr>
<td>3.3</td>
<td>The predicted correct (solid lines) and incorrect (dashed lines) joint response time densities based on the racing diffusion model given extremely high threshold variability for errors.</td>
<td>70</td>
</tr>
</tbody>
</table>
3.4 The decision (left-hand side) and interference (right-hand side) processes in both the congruent (top) and incongruent (bottom) condition. The parameter $\kappa$ is the threshold, $\xi$ refers to the drift rate, and $\rho$ governs the frequency in which a response is generated from a particular process. The arrows represent the average path of the evidence accumulation, and their color indicates the choice to which they correspond.

5.1 Quantile-probability plot showing circles and crosses (representing the congruent and incongruent conditions, respectively) for the observed 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles averaged over subjects on the y-axis and probability of a correct or incorrect response on the x-axis. Ellipses represent plus or minus two standard errors.

5.2 Quantile-probability plot showing circles and crosses (representing the payoffs and no payoffs conditions, respectively) for the observed 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles averaged over subjects on the y-axis and probability of a correct or incorrect response on the x-axis. Ellipses represent plus or minus two standard errors.

5.3 The left plot shows changes in average accuracy by block, while the right plot shows changes in the averaged 0.1, 0.3, 0.5, 0.7, and 0.9 quantiles by block. Shaded regions denote two standard errors above and below the mean, providing approximate 95% confidence intervals. The dashed lines represent the average accuracy (on the left) and average median response time (on the right) across all blocks.

5.4 Averages for two different measures of post-error slowing. The traditional measure is the mean of correct response times following an error minus the mean of correct response times following a correct choice. The robust measure is the mean of the difference in response times for trials following an error and trials preceding an error. Lines represent two standard errors above and below the mean.

5.5 Example showing the visual inspection process used to establish the burn in period and identify poorly performing chains. The solid vertical line denotes the chosen burn-in period. The dashed red lines denote the interval used to identify poorly performing chains. Chains that lie outside this interval for over 100 iterations were removed.
5.6 Estimates of the WAIC for pooled observations across all subjects for each model variant. Error bars represent two standard errors above and below each value. ................................................. 133

5.7 Quantile-probability plots of correct and error responses per congruency condition for select subjects. Ellipses are the 95% credible intervals for model predictions. ................................................. 135

5.8 Accuracy per block and session for select subjects. Gray lines represent the median and lower and upper limits of the 95% credible intervals for the model predictions. ................................................. 138

5.9 The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles by block for select subjects. Gray bands represent the 95% credible intervals for model predictions regarding the median (dark gray), 0.3 and 0.7 (medium gray) and 0.1 and 0.9 (light gray) quantiles. ................................................. 140

5.10 Quantile-probability plots for each model variant showing the 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles for Subject 12, based on whether the preceding error was fast or slow. ......................... 142

5.11 The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles for trials following an error per each subject. Gray boxes denotes the 95% credible intervals for model predictions, with the median represented in dark gray. ................................................. 143

5.12 Histograms of the summed scores for the “Doubts about actions”, “Concerns over mistakes”, and STAI subscales by session. ......................... 145

5.13 Credible intervals (95%, 90%, 80%, and 50%) and median estimates for the intercept term of each point estimate from the cognitive model. 147

5.14 Credible intervals (95%, 90%, 80%, and 50%) and median estimates for the “Concerns over mistakes” term of each point estimate from the cognitive model. ................................................. 149

5.15 Credible intervals (95%, 90%, 80%, and 50%) and median estimates for the “Doubts about actions” term of each point estimate from the cognitive model. ................................................. 151
5.16 Credible intervals (95%, 90%, 80%, and 50%) and median estimates for the STAI term of each point estimate from the cognitive model.  

5.17 Quantile-probability plots of correct and error response times for trials following an error for select subjects. Ellipses are the 95% credible intervals for model predictions.  

A.1 Quantile-probability plots of correct and error responses per congruency condition for Subject 1, 9, and 11. Ellipses are the 95% credible intervals for model predictions. The model displayed good fit to the data.  

A.2 Quantile-probability plots of correct and error responses per congruency condition for Subject 14, 19, and 24. Ellipses are the 95% credible intervals for model predictions. The model displayed good fit to the data.  

A.3 Quantile-probability plots of correct and error responses per congruency condition for Subject 26, 27, and 29. Ellipses are the 95% credible intervals for model predictions. The model displayed good fit to the data. 

A.4 Quantile-probability plots of correct and error responses per congruency condition for Subject 2, 3, 8, and 10. Ellipses are the 95% credible intervals for model predictions. The model displayed adequate fit to the data.  

A.5 Quantile-probability plots of correct and error responses per congruency condition for Subject 14, 19, and 24. Ellipses are the 95% credible intervals for model predictions. The model displayed adequate fit to the data.  

A.6 Quantile-probability plots of correct and error responses per congruency condition for Subject 4, 5, 6, and 7. Ellipses are the 95% credible intervals for model predictions. The model displayed poor fit to the data.  

A.7 Quantile-probability plots of correct and error responses per congruency condition for Subject 13, 15, and 16. Ellipses are the 95% credible intervals for model predictions. The model displayed poor fit to the data.
A.8 Quantile-probability plots of correct and error responses per congruency condition for Subject 17, 18, and 20. Ellipses are the 95% credible intervals for model predictions. The model displayed poor fit to the data. ............................................................. 190

A.9 Quantile-probability plots of correct and error responses per congruency condition for Subject 17, 18, and 20. Ellipses are the 95% credible intervals for model predictions. The model displayed poor fit to the data. ............................................................. 191

A.10 Quantile-probability plots of correct and error responses per congruency condition for Subject 21, 23, and 29. Ellipses are the 95% credible intervals for model predictions. The model displayed poor fit to the data. ............................................................. 192

A.11 Accuracy per block and session for Subject 1, 2, 3, and 4. Gray lines represent the median and lower and upper limits of the 95% credible intervals for the model predictions. .................................................. 193

A.12 Accuracy per block and session for Subject 5, 6, 7, and 8. Gray lines represent the median and lower and upper limits of the 95% credible intervals for the model predictions. .................................................. 194

A.13 Accuracy per block and session for Subject 9, 10, 11, and 12. Gray lines represent the median and lower and upper limits of the 95% credible intervals for the model predictions. .................................................. 195

A.14 Accuracy per block and session for Subject 13, 14, 15, and 16. Gray lines represent the median and lower and upper limits of the 95% credible intervals for the model predictions. .................................................. 196

A.15 Accuracy per block and session for Subject 17, 18, 19, and 20. Gray lines represent the median and lower and upper limits of the 95% credible intervals for the model predictions. .................................................. 197

A.16 Accuracy per block and session for Subject 21, 22, and 23. Gray lines represent the median and lower and upper limits of the 95% credible intervals for the model predictions. .................................................. 198
A.17 Accuracy per block and session for Subject 24, 25, and 26. Gray lines represent the median and lower and upper limits of the 95% credible intervals for the model predictions. .............................. 199

A.18 Accuracy per block and session for Subject 27, 28, and 29. Gray lines represent the median and lower and upper limits of the 95% credible intervals for the model predictions. ..................... 200

A.19 The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles by block for Subject 1, 2, 3, and 4. Gray bands represent the 95% credible intervals for model predictions regarding the median (dark gray), 0.3 and 0.7 (medium gray) and 0.1 and 0.9 (light gray) quantiles. ............... 201

A.20 The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles by block for Subject 5, 6, 7, and 8. Gray bands represent the 95% credible intervals for model predictions regarding the median (dark gray), 0.3 and 0.7 (medium gray) and 0.1 and 0.9 (light gray) quantiles. ............... 202

A.21 The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles by block for Subject 9, 10, 11, and 12. Gray bands represent the 95% credible intervals for model predictions regarding the median (dark gray), 0.3 and 0.7 (medium gray) and 0.1 and 0.9 (light gray) quantiles. .......... 203

A.22 The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles by block for Subject 13, 14, 15, and 16. Gray bands represent the 95% credible intervals for model predictions regarding the median (dark gray), 0.3 and 0.7 (medium gray) and 0.1 and 0.9 (light gray) quantiles. .......... 204

A.23 The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles by block for Subject 13, 14, 15, and 16. Gray bands represent the 95% credible intervals for model predictions regarding the median (dark gray), 0.3 and 0.7 (medium gray) and 0.1 and 0.9 (light gray) quantiles. .......... 205

A.24 The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles by block for Subject 17, 18, 19, and 20. Gray bands represent the 95% credible intervals for model predictions regarding the median (dark gray), 0.3 and 0.7 (medium gray) and 0.1 and 0.9 (light gray) quantiles. .......... 206
A.25 The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles by block for Subject 21, 22, and 23. Gray bands represent the 95% credible intervals for model predictions regarding the median (dark gray), 0.3 and 0.7 (medium gray) and 0.1 and 0.9 (light gray) quantiles. . . . . . . . . . 207

A.26 The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles by block for Subject 24, 25, and 26. Gray bands represent the 95% credible intervals for model predictions regarding the median (dark gray), 0.3 and 0.7 (medium gray) and 0.1 and 0.9 (light gray) quantiles. . . . . . . . . . 208

A.27 The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles by block for Subject 27, 28, and 29. Gray bands represent the 95% credible intervals for model predictions regarding the median (dark gray), 0.3 and 0.7 (medium gray) and 0.1 and 0.9 (light gray) quantiles. . . . . . . . . . 209

A.28 Quantile-probability plots of correct and error response times for trials following an error for Subject 1, 2, 3, and 4. Ellipses are the 95% credible intervals for model predictions. . . . . . . . . . . . . . . . . . 210

A.29 Quantile-probability plots of correct and error response times for trials following an error for Subject 5, 6, 7, and 8. Ellipses are the 95% credible intervals for model predictions. . . . . . . . . . . . . . . . . . 211

A.30 Quantile-probability plots of correct and error response times for trials following an error for Subject 9, 10, 11, and 12. Ellipses are the 95% credible intervals for model predictions. . . . . . . . . . . . . . . . . . 212

A.31 Quantile-probability plots of correct and error response times for trials following an error for Subject 13, 14, 15, and 16. Ellipses are the 95% credible intervals for model predictions. . . . . . . . . . . . . . . . . . 213

A.32 Quantile-probability plots of correct and error response times for trials following an error for Subject 17, 18, 19, and 20. Ellipses are the 95% credible intervals for model predictions. . . . . . . . . . . . . . . . . . 214

A.33 Quantile-probability plots of correct and error response times for trials following an error for Subject 21, 22, and 23. Ellipses are the 95% credible intervals for model predictions. . . . . . . . . . . . . . . . . . 215
A.34 Quantile-probability plots of correct and error response times for trials following an error for Subject 24, 25, and 26. Ellipses are the 95% credible intervals for model predictions. .............................................. 216

A.35 Quantile-probability plots of correct and error response times for trials following an error for Subject 27, 28, and 29. Ellipses are the 95% credible intervals for model predictions. .............................................. 217
Chapter 1: Introduction

In clinical psychology, perfectionism has long been believed to play an important role in the cause, maintenance, and course of several psychopathological states (Shafran, Cooper, & Fairburn, 2002; Shafran & Mansell, 2001). One long standing conception that has proven useful is that perfectionism is the setting of “high standards of performance which are accompanied by tendencies for overly critical evaluations of one’s own behavior” (Frost, Marten, Lahart, & Rosenblate, 1990, p. 450). However, perfectionism remains an ill-defined and poorly understood construct (Egan, Wade, & Shafran, 2011; Shafran et al., 2002), and it is only recently that the cognitive and neural aspects of perfectionism have come under study.

Because of its involvement in a wide variety of disorders, and the stress and anxiety it can cause in the daily life of non-clinical populations such as university students, understanding how perfectionism comes about and influences behavior is of great importance. Perfectionism has been shown to influence a wide variety of behaviors, including decision making (e.g. Boisseau, Thompson-Brenner, Pratt, Farchione, & Barlow, 2013; Brand & Alstötter-Gleich, 2008). Therefore, to expand and further clarify the relationship between decision-making and perfectionism, the goal of this project is to test hypotheses about how perfectionism influences specific mechanisms
of simple choice in a forced-choice two-alternative task. I will use mathematical models of decision making to better explain the cognitive processes at work. Such models have proven very useful in separating out the elementary components of decisions for a wide array of tasks, and can identify and point to causes for abnormal cognitive processes (e.g. White, Ratcliff, Vasey, & McKoon, 2010b).

I first discuss the conception of perfectionism that I will rely on and its importance for several clinical disorders. Next, I elaborate the current status of the research on the cognitive underpinnings for this construct, with a particular focus on post-error behavior and cognitive control. I note several caveats in using post-error behavior to measure cognitive control, including issues such as an over-reliance on aggregation, a disregard for multiple types of post-error adjustments, and the risks of assuming post-error slowing represents more cautious responding. I then discuss how mathematical models of simple choice and response time may solve such issues, in addition to the other more general benefits of a modeling approach.

In the following I explain my experimental method, discuss the structure of the model I use to fit my data, the analyses I conducted to fit and evaluate the model, and the results. I end with a discussion of the implications my project has for future research seeking to examine the underlying cognitive mechanisms of perfectionism. My project adds to the current literature by introducing a novel model for the Simon task and by providing a direct test of a core assumption regarding post-error behavior and perfectionism that past researchers have failed to check. Furthermore, my project is one of the first to examine the link between perfectionism and post-error behavior via a principled mathematical model.
1.1 Perfectionism

Interest in perfectionism has rapidly increased in the past three decades. For instance, Flett and Hewitt (2002) noted that in the 1980s, only 102 publications had the term “perfectionism” as a keyword, but the number had increased to 336 by the 1990s. Now, a keyword search can produce over a thousand articles. Despite this increase in interest, the definitions, dimensions, and nature of perfectionism are still elusive (Egan et al., 2011). Nonetheless, a large body of research has found that perfectionism has a noticeable maladaptive role in many clinical disorders (Shafran & Mansell, 2001). In the following, I briefly review the way that perfectionism has been conceptualized and measured, some of the findings and links between perfectionism and psychopathology, and the work examining the neural basis of perfectionism and its relation to simple decision making and post-error slowing.

1.1.1 Conceptualization of Perfectionism

At first, perfectionism appears simple to define. There is a great deal of agreement among researchers that it reflects the setting of excessively high standards for performance (Burns, 1980; Flett & Hewitt, 2002; Frost et al., 1990; Hamachek, 1978; Shafran & Mansell, 2001). However, as the research and theories on perfectionism have grown, there are now a large and varied number of conceptualizations and definitions (Flett & Hewitt, 2002). For instance, Flett and Hewitt note that perfectionism can be viewed as maladaptive or beneficial, as other- or self-oriented, or as active or passive, to name a few examples. A wide variety of inventories have been devised, each examining different aspects of perfectionism, such as the Frost Multidimensional
Perfectionism Scale (FMPS; Frost et al., 1990), the Hewitt and Flett Multidimensional Perfectionism Scale (HMPS; Hewitt & Flett, 1991), the Almost Perfect Scale (Slaney, Ashby, & Trippi, 1995), and the Perfectionism Cognitions Inventory (Flett, Hewitt, Blankstein, & Gray, 1998). This diversity of scales highlights the varied viewpoints on perfectionism and what it is. There appears to be general consensus that perfectionism is a multidimensional construct (e.g. Hewitt & Flett, 1991; Frost et al., 1990), as two of the most predominantly used inventories, the FMPS and HMPS, both define perfectionism as multidimensional, albeit in different ways (Egan et al., 2011). For my project, I will focus on the conceptualization of perfectionism put forth by Frost et al. (1990) and measured in the FMPS, as it is the inventory predominantly used in the examination of perfectionism and post-error slowing.

**The FMPS.** In deriving the FMPS, Frost and his collaborators initially identified two main components of perfectionism that were deemed important based on a review of past literature: personal standards and self-evaluation. Frost et al. also further elaborated on several kinds of critical evaluations, resulting in five features they deemed important.

- Frost et al. noted that a consistent theme in past work defining perfectionism was the setting of excessively high personal standards of performance. These high standards, when accompanied by tendencies for overly critical evaluations of one’s own behavior, lead to maladaptive behaviors.

- Frost et al. also noted that perfectionists appear to be overly concerned with mistakes, such that even minor ones result in the perception that standards have not been met. These concerns are believed to be exacerbated by dichotomous
thinking, where any flaw that a person perceives in his or her performance is interpreted as failure, leaving no margin of error.

- Another feature of self-evaluation that Frost et al. identified was that perfectionists tend to have vague doubts about the quality of their work, a sense that a job is not satisfactorily completed.

- Additionally, Frost et al. noted the importance of the link between the self-evaluations of perfectionists and assumptions about parental expectations and approval or disapproval.

- Finally, Frost et al. noted that perfectionists were often characterized as over-emphasizing precision, order, and organization.

Frost and his colleagues designed the FMPS on the basis of these five features. They drew items from existing measures of perfectionism and obsessions, and created several new items. Reliability analyses and two factor analyses conducted in samples of 232 and 178 female undergraduates respectively resulted in the final inventory, consisting of 35 items with six subscales. The subscales measure six factors: “Concerns over mistakes” (nine items, e.g., “I should be upset if I make a mistake”), “Organization” (six items, e.g., “Neatness is very important to me”), “Parental criticism” (four items, e.g., “I never felt like I could meet my parents’ standards”), “Parental expectations” (five items, e.g., “My parents wanted me to be the best at everything”), “Personal standards” (seven items, e.g., “I have extremely high goals”) and “Doubts about actions” (DA; four items, e.g., “Even when I do something very carefully, I often feel that it is not quite right”). Frost et al. provided internal consistencies,
reporting a Cronbach’s alpha that ranged from 0.77 to 0.93. The authors also found that the FMPS showed large correlations with other perfectionism measures.

A large number of studies examining various psychopathological symptoms and diagnostic groups have relied on the FMPS (Flett & Hewitt, 2002). However, it is important to note that the FMPS has several issues. For example, the factor structure of the inventory is highly unstable, with studies also finding three-factor (Purdon, Antony, & Swinson, 1999), four-factor (Khawaja & Armstrong, 2005; Stöber, 1998; Stumpf & Parker, 2000), and five-factor (Cox, Enns, & Clara, 2002; Stallman & Hurst, 2011) solutions. Indeed, in the original development of the scale, Frost et al. found that several items would change which factors they loaded onto from one sample of subjects to the next.

Furthermore, items in the FMPS have also shown high loadings on multiple factors at once, complicating the ability of the inventory to provide straightforward interpretations of the underlying construct of perfectionism (Cox et al., 2002; Stallman & Hurst, 2011). Stallman and Hurst (2011) recommend the exclusion of six items to improve stability of the scale, while Cox and his colleagues recommended the removal of nine. Nonetheless, inventories like the FMPS have provided a convenient way to conceptualize perfectionism, and have been a useful tool to further study this construct. Indeed, such inventories have provided important evidence in the widespread involvement of perfectionism in the etiology and maintenance of many clinical disorders.
1.1.2 Clinical Relevance

Perfectionism has been identified as playing a key role in a wide variety of disorders, including eating disorders, depression, and obsessive-compulsive personality disorder (Shafran et al., 2002; Egan et al., 2011). It has also been implicated in hoarding disorder (e.g. Frost & Gross, 1993; Frost & Hartl, 1996). Given the distress, impairment in social and occupational life, and in some cases fatal consequences that these disorders can cause, it is important to better understand the mechanisms of perfectionism.

Eating disorders. Perfectionism has been implicated as an underlying cause of eating disorders such as anorexia nervosa. Perfectionist attitudes, among other things, are hypothesized to drive a person’s need for self control, eventually resulting in restrictive and harmful eating practices (Fairburn, Shafran, & Cooper, 1999). Indeed, several large-scale community studies and systematic reviews have found perfectionism to be a specific risk factor for patients with anorexia nervosa, bulimia nervosa, and binge eating disorder (e.g. Bardone-Cone et al., 2007; Fairburn, Cooper, Doll, & Welch, 1999). Perfectionism’s role in eating disorders is further supported by the finding that elevated levels of perfectionism are found even in samples of weight restored anorexia nervosa and bulimia nervosa patients (Bastiani, Rao, Weltzin, & Kaye, 1995; Kaye et al., 2004).

As noted by Shafran and Mansell (2001), many researchers believe perfectionism to be part of the phenomenology of anorexia nervosa, and as such several inventories for eating disorders include a perfectionism rating scale, such as the Eating Disorders
Inventory (Garner, Olmstead, & Polivy, 1983), the Neurotic Perfectionism Questionnaire (Mitzman, Slade, & Dewey, 1994), and the Setting Conditions for Anorexia Nervosa Scale (Slade & Dewey, 1986). In both student and psychiatric populations, several studies have found associations between higher perfectionism scores and eating disorder symptoms (e.g. Bastiani et al., 1995; Sassaroli et al., 2008).

The outcome of eating disorders, in particular for anorexia nervosa, is often poor. For instance, a meta-analysis by Arcelus, Mitchell, Wales, and Nielsen (2011) notes that only 46% of anorexia nervosa patients fully recovered, whereas 20% remained chronically ill for the long term. They also found that anorexia nervosa had the highest mortality rate, with a weighted annual mortality rate of 5 per 1000 person-years (the product of the number of years times the number of members of the population of people suffering from anorexia nervosa). A better understanding of perfectionism, given its role in the development of eating disorders, may help improve treatment and recovery outcome.

**Depression.** Several studies and case studies have also found links between perfectionism and depression. For instance, Burns and Beck (1978), examining the case of a young woman who experienced unremitting depression and bouts of self-harm, concluded that her problem arose from exceedingly perfectionistic, all-or-none standards. She developed depressive symptoms whenever she perceived her performance as being less than perfect, resulting in decreased productivity that exacerbated her negative evaluation, resulting in a vicious cycle of depressive thoughts and behaviors.

Frost et al. (1990) found that several subscales of the FMPS correlated significantly with the dependency depression and self-critical depression subscales of the
Depressive Experiences Questionnaire (DEQ; Blatt, Quinlan, Chevron, McDonald, & Zuroff, 1982). The study by Sassaroli et al. (2008) also found that a sample of clinically depressed individuals scored significantly higher on the “Concerns over mistakes” subscale of the FMPS compared to controls. A study by Enns and Cox (1999) administered a large battery of inventories, including the FMPS, the HMPS, the NEO-five factor inventory (Costa & McCrae, 1992), the DEQ, the Beck Depression Inventory (BDI; Beck, Steer, & Garbin, 1988), and the Hamilton Depression Ratings (Hamilton, 1960). The authors found several correlations across the many subscales for the inventories. After partialling out variance shared with neuroticism and extraversion (the two personality factors possessing significant correlations with the perfectionism measures), the subscales of perfectionism reflecting concern over mistakes and socially prescribed perfectionism still correlated with the cognitive distortions subscale of the BDI, and the “Concerns over mistakes” subscale still correlated with the Hamilton Depression Ratings.

Perfectionism has important implications for the negative consequences of depression and its treatment. For example, Egan et al. (2011) note that several studies have found that, in student and psychiatric samples, increased suicidal ideation is associated with higher scores on several subscales of perfectionism, such as socially prescribed perfectionism, concern over mistakes, and doubts about actions. Furthermore, in a reanalysis of a large-scale investigation of the effects of four brief treatments (Blatt, Quinlan, Pilkonis, & Shea, 1995), higher levels of perfectionism predicted negative treatment outcome regardless of whether the treatment involved cognitive behavioral therapy or pharmacotherapy.
Anxiety. As noted by Egan et al. (2011), perfectionism has shown a robust correlation with anxiety. For instance, a study by Wirtz et al. (2007) examined cortisol levels in 50 middle age men who underwent a psychosocial stress test which combines a speech in front of an audience and a mental arithmetic task. The authors also administered several questionnaires, including the FMPS. The results indicated that the perfectionism subscale combining concern over mistakes and doubts about actions correlated significantly with cortisol levels \( r = 0.32, p < 0.05 \). In an exploratory stepwise regression, even after controlling for age, body mass index, mean arterial pressure, and other personality factors like neuroticism and extraversion, perfectionism was still a significant predictor of cortisol levels, uniquely accounting for 18% of the variance. Moreover, links between perfectionism and an assortment of anxiety disorders including obsessive-compulsive disorder, social anxiety, and panic disorder have been found by several studies (Egan et al., 2011).

Obsessive compulsive disorder. The Obsessive Compulsive Cognitions Working Group (1997) identified perfectionism as one of one six key belief domains in the etiology and maintenance of obsessive compulsive disorder. A common report of individual case studies is that patients suffering from obsessive compulsive disorder see imperfections everywhere and attempt to cope by engaging in compulsive behaviors (Frost, Novara, & Rhéaume, 2002). Moreover, clinical samples of individuals with obsessive compulsive disorder have significantly elevated perfectionism for several different inventories (e.g. Egan et al., 2011; Sassaroli et al., 2008). A study by Chik, Whittal, and O’Neill (2008) found that higher scores on the “Doubts about actions” subscale uniquely predicted worse treatment outcome for a sample of 118 patients.
who underwent cognitive therapy or exposure and response prevention for obsessive compulsive disorder. On a related note to obsessive compulsive disorder, perfectionism has also been implicated in compulsive hoarding. For instance, Frost and Gross (1993) found significant correlations between the FMPS and an inventory assessing hoarding, and Frost and Hartl (1996) posited that one cognitive basis underlying compulsive hoarding is elevated concern over making mistakes.

The brief review given above provides only a few examples of the large number of associations found between perfectionism and psychopathology. These associations have led researchers such as Egan et al. (2011) to suggest that that perfectionism may be a transdiagnostic process (a process that occurs across multiple disorders, and is both a risk factor and a maintaining mechanism). However, while there has been a great deal of research relating perfectionism to a variety of disorders, there has also been great variability in the results. For instance, Enns and Cox (1999) argue that the most noteworthy observation in their study was the great variability in the magnitude (and direction) of correlations between the perfectionism dimensions and depression symptoms.

A more precise definition and better understanding of the underlying cognitive mechanisms of perfectionism may aid in addressing this issue of variability. One method that may aid answering such questions and in providing a more concrete definition of perfectionism is to examine how perfectionism relates to specific behaviors like decision making. For instance, the FMPS has been used in several studies examining whether scale scores on perfectionism can predict decision making performance (Boisseau et al., 2013; Brand & Altstötter-Gleich, 2008; Pieters et al., 2007; Schrijvers, De Bruijn, Destoop, Hulstijn, & Sabbe, 2010; Tops, Koole, & Wijers, 2013).
Furthermore, the studies by Pieters et al. (2007), Schrijvers et al. (2010), and Tops et al. (2013) have sought to explore the neural basis of perfectionism.

1.1.3 Neural Basis

Recently, electroencephalography (EEG) studies have examined the impact of individual differences in perfectionism on various event-related potentials (ERPs). The studies in particular have focused on two ERPs, the “Error related negativity” (ERN; Gehring, Goss, Coles, Meyer, & Donchin, 1993) also known as “Error Negativity” (Ne; Falkenstein, Hohnsbein, Hoormann, & Blanke, 1991), and the “Error Positivity” (Pe; Falkenstein et al., 1991), as observed during the Eriksen Flanker task (Eriksen & Eriksen, 1974). The studies examining perfectionism used a version of the task that has a subject make a different response based on a central letter (either H or S). The central letter can be flanked either by congruent letter strings (e.g. HHHHHH) or incongruent letter strings (e.g. SSHSSS). The flanker task can also use arrows in which the flankers either point in the same or opposing direction (Tops et al., 2013).

The ERN

Many EEG studies rely on the flanker task since it reliably generates the ERN. Independently found by Falkenstein et al. (1991) and Gehring et al. (1993), the ERN is a response-locked potential that appears as a negative deflection peaking approximately 50-150 milliseconds after subjects make an error. As noted by Holroyd and Coles (2002), its spatial distribution lies over frontal–central regions of the scalp, reaching maximum amplitude in a region over the supplementary motor area. Gehring et al. (1993) showed that the amplitude of the ERN increases when subjects are more motivated by payoffs to strive for accuracy, and Bernstein, Scheffers, and Coles (1995)
demonstrated that when subjects could make two different types of response errors (a response with the wrong hand or wrong finger), the amplitude of the ERN is highest when a combination of the two errors was committed, suggesting that the ERN is sensitive to the degree of error.

The ERN is believed to reflect activity in the anterior cingulate cortex (ACC). Dehaene, Posner, and Tucker (1994), using the brain electric source analysis technique for source localization found evidence that the ERN is generated within that region. A study by Debener et al. (2005) carried out simultaneous recording of EEG and functional Magnetic Resonance Imaging (fMRI) data, and found that error related negativity of the EEG predicted fMRI activity in the rostral cingulate zone (a region of the ACC involved more with cognitive functions). Note that some recent work has challenged the theory that the ACC is the source of the ERN, and that the posterior cingulate cortex may instead be involved (Agam et al., 2011).

The ERN has been hypothesized to reflect some form of action and performance monitoring. For instance, Holroyd and Coles (2002) posit that the ERN is generated due to the mesencephalic dopamine system conveying negative reinforcement learning signals that disinhibit the apical dendrites of motor neurons in the ACC. In other words, when actual outcomes are worse than the expected or predicted outcomes, greater cognitive control is recruited. Another important theory accounting for the ERN is conflict monitoring (e.g. Botvinick, Braver, Barch, Carter, & Cohen, 2001; Botvinick, Cohen, & Carter, 2004). In this theory, the dorsal ACC responds to the occurrence of conflicts in information processing (such as response competition), triggering strategic adjustments in cognitive control. The ERN is a result of transient activation of the ACC in association with the commission of errors (believed to occur
because of belated activation of the correct response during the commission of the error, resulting in conflict).

A more recent theory is the Predicted Response Outcome model (J. W. Brown, 2013). In this model, the ACC learns to predict the consequences of various outcomes, errors in particular. In addition, as posited by Holroyd and Coles (2002), the ACC compares the actual and predicted outcomes. One of the predictions from the computational model is the generation of an ERN to infrequent errors. Despite the differing theories, all involve performance monitoring and some reaction to the commission of errors, which is posited to be elevated in perfectionist individuals. Hence, the ERN (and indeed, any measure of performance adjustment), is a logical candidate to examine to better understand perfectionism.

**The Pe**

Another ERP related to perfectionism appears to be the Pe (Falkenstein et al., 1991). The Pe is a slow positive wave with a diffuse scalp distribution and a maximum amplitude between 200 to 400 ms, observed after the occurrence of the ERN. The Pe has not been as extensively studied as the ERN, but it is hypothesized that this component reflects greater conscious awareness of error commissions and subjective, affective evaluation of the error (Falkenstein, Hoormann, Christ, & Hohnsbein, 2000; Overbeek, Nieuwenhuis, & Ridderinkhof, 2005). For instance, Falkenstein et al. (2000) found that subjects who made many errors exhibited a smaller amplitude for the Pe compared to subjects who made fewer errors, which the authors suggested indicates that subjects who made more errors care less. However, Overbeek et al. (2005) notes that other studies have failed to replicate this result, and a study by Hajcak,
McDonald, and Simons (2004) found that subjects scoring high on negative affect in fact have been reported to exhibit a smaller Pe than low negative affect subjects.

There is more support for the Pe reflecting greater awareness of errors. A study by Nieuwenhuis, Ridderinkhof, Blom, Band, and Kok (2001) found that in an antisaccade task, in which subjects had to judge the accuracy of the responses on each trial, subjects only noticed about half of the 20% or so erroneous responses were in fact incorrect. The Pe for the perceived errors was larger compared to that for the unperceived errors. Studies by Endrass, Franke, and Kathmann (2005) and Hewig, Coles, Trippe, Hecht, and Miltner (2011) have replicated the finding of a larger Pe for perceived errors. It is worth mentioning that the ERN also appears to still be impacted by conscious awareness, albeit less so than the Pe. Nonetheless, the Pe has provided an additional neural marker by which to assess perfectionism.

**EEG Studies of Perfectionism**

As mentioned before, three different students have relied on EEG methods to gain insight into the cognitive mechanisms underlying perfectionism.

**Perfectionism, the ERN, and eating disorders.** One of the initial examinations of perfectionism using EEG methods was the study by Pieters et al. (2007). The researchers gathered behavioral and EEG data from seventeen underweight anorexia nervosa patients and nineteen matched healthy controls on a flanker task using letter strings (e.g. HHSHH). The goal of the study was to examine differences in action monitoring between the anorexia patients and controls. Importantly, the authors also administered the FMPS to assess perfectionism.
The subjects completed six blocks of 100 trials each, where for each block half the trials had congruent stimuli while the other half were incongruent. To avoid the influence of accuracy on the ERN amplitude, individual reaction time deadlines were calculated during an initial practice block of 60 trials. After the practice block, the deadline for a subject was determined by adding 0.5 standard deviations to the average reaction time. This procedure ensured an accuracy of about 75 - 80% for each subject. Subjects received feedback indicating whether they were correct, incorrect, or too slow. Equal emphasis was placed on speed and accuracy.

For each subject, the ERPs for correct and incorrect responses were averaged, time-locked to response onset. Using averaged ERPs for incorrect responses for each subject, the ERN was determined by subtracting the most positive peak between 80 ms before and after response onset from the most negative peak between 0 to 200 ms after response onset at electrode Cz (located in the center at the top of the scalp). The Pe was determined as the mean amplitude of the difference wave (incorrect minus correct response) in the 200 to 400 ms time window following response onset at electrodes Fz, FCz, Cz, and Pz (i.e. midline locations along the scalp).

Several statistical analyses were then carried out to test the hypotheses of the study. The authors entered averages for response times and amplitudes in a repeated measures ANOVA with group (controls versus anorexia nervosa patients) as a between-subjects factor, while congruency and correctness were within-subject factors. When deemed significant, error rates were included as a covariate. The researchers also examined post-error slowing, quantified as the difference between response times for correct responses preceded by an incorrect response and correct responses preceded by a correct response. Separate t-tests for each group were then
carried out to test for differences. Finally, ERN amplitudes and the proportion of errors were correlated with the overall FMPS score.

The authors found that more errors were made by the control group, and the control group displayed a significant amount of post-error slowing, while the anorexia nervosa patients did not. In fact, the latter group was much less likely to commit a second error following an incorrect response (4.73%) compared to the control group (11.27%). When including error rates as a covariate, the ERN was significantly smaller for anorexia nervosa patients compared to controls. For both groups, the ERN exhibited a significant negative correlation with error rates, and the ERN correlated significantly with overall FMPS scores for controls, but not for anorexia patients. No significant relationships were found for the Pe.

Pieters et al. note that their results indicate that anorexia patients do engage in a more controlled, perfectionistic response style. They posit that the smaller ERN and lack of a correlation between the ERN and FMPS for anorexia patients may reflect the blunted cingulate activity found for this group. At the very least, the correlation between the ERN and FMPS and the significant amount of post-error slowing found for the control group provides an important starting place for examining the relation between post-error slowing and perfectionism.

**Perfectionism, the Pe, and depression.** Schrijvers et al. (2010) examined relations between the ERN, Pe, and perfectionism using a sample of patients suffering severe depression. The authors examined the hypothesis that depressed patients with higher levels of subjective perfectionism and trait anxiety would generate larger ERN amplitudes compared to those reporting lower levels. 39 patients with major
depressive disorder completed a Dutch version of the FMPS and the Trait form of the State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, & Lushene, 1970). The researchers gathered behavioral and EEG data on a flanker task using the same experimental design as Pieters et al. (2007).

The ERN was calculated in a similar manner as done by Pieters et al., but this time for electrodes Cz and Fz. The Pe was also calculated in a similar manner as Pieters et al., but only across electrodes Cz and Pz. The authors used repeated measures ANOVA and regression analyses to assess the significance and relations between the amplitude of the ERN and Pe components and behavioral variables. Furthermore, bivariate correlations were carried out to examine the relationship between the amplitudes and the overall and subscale scores for the two inventories. As is typical for the flanker task, Schrijvers et al. found significant post-error slowing. Moreover, the “Concerns over mistakes” subscale of the FMPS significantly correlated with the amplitude of the Pe, while the “Doubts about actions” subscale correlated with the ERN amplitude.

The authors argue that the significant correlation between the Pe and the “Concerns over mistakes” subscale fit well with the hypothesis that the Pe reflects an emotional error-assessment process, given that the “Concerns over mistakes” subscale indicates negative reactions to mistakes, interpreting them as failures that will lead to a loss of respect. However, no correlations between the inventory scores and any of the behavioral variables (e.g. post-error slowing) were found. Schrijvers et al. suggest that, because the flanker task is a perceptually and conceptually straightforward task, there may be insufficient demands on processing resources for the effects of perfectionism and depression to be detectable in the behavioral measures.
Perfectionism, the Pe, and the “Concerns over mistakes” subscale. One of the most recent studies examining perfectionism using EEG methods was by Tops et al. (2013). The authors further examined the relation between the “Concerns over mistakes” subscale of the FMPS and the ERN and Pe components. They posited that the lack of correlation between the ERN amplitude and scores for the “Concerns over mistakes” subscale indicates that the subscale is not associated with performance monitoring but instead indicates increased rumination and emotion after an error. Hence, they sought to replicate the association between the “Concerns over mistakes” subscale and the Pe.

However, Tops et al. noted a temporal distinction between an early Pe occurring 200 to 300 ms following a response (which does not discriminate between aware and unaware errors) and a late Pe occurring 400 to 600 ms following a response. The late Pe is believed to represent activity in the inferior frontal gyrus and anterior insula. The authors hypothesize that this later Pe is therefore more likely to be related to concern over mistakes. The authors had 16 female subjects complete the “Concerns over mistakes” subscale and then finish 13 blocks of 60 trials each of the flanker task (where the first three blocks were practice), this time using arrows as the stimuli (e.g. <<><>). As was done in the previous two studies, individual response deadlines were estimated based on the first practice block, using the mean response time for correct responses plus one standard deviation as the deadline. If subjects were too slow, the phrase ‘late’ was given as feedback. Otherwise, subjects saw a photograph tinted green for correct responses or red for incorrect. To heighten emotional salience, the photograph had a person’s face with either a happy or disgusted expression. The presented expression was randomly selected.
Mean amplitudes for the ERN (identified visually as a negative peak at 100 ms for electrode Cz) were calculated for each subject within the time window of 88 to 112 ms. The average amplitude for the early Pe was calculated within the time window of 150 to 350 ms following the response at electrode Pz. The average amplitude for the late Pe was calculated in the time window of 400 to 500 ms at electrode Fz. Tops et al. found that scores on the “Concerns over mistakes” subscale did not predict response times. A general linear model with flanker congruity as a within subjects factor, “Concerns over mistakes” scores as a predictor, and the proportion of correct responses as the dependent variable, found no main effect or interaction with the “Concerns over mistakes” subscale. No effect of concern over mistakes was found on any of the ERN or for the early Pe.

The authors speculate that the feedback given trial to trial may have attenuated performance monitoring. This may have decreased excessive error monitoring related to anxiety traits, or increased anticipatory processes related to potentially distressing feedback. The late Pe was significantly modulated by whether a response was correct and by concern over mistakes, as at electrode Fz for errors subjects scoring higher on the “Concerns over mistakes” subscale had more positive amplitudes.

Tops et al. hypothesized that high concern over mistakes relates to a decreased negative-going slope in amplitude relative to the feedback. The authors suggest that this may reflect the stimulus-preceding negativity, interpreted to be anticipation of impending stimuli. Individuals who score highly on “Concerns over mistakes” may downregulate distress over the ambiguous social feedback (since a disgusted face could occur for both correct and incorrect responses) through anticipatory avoidance of emotional processing for negative feedback. Therefore, the researchers carried out
several exploratory studies to further examine the late positive potential (associated with emotional regulation).

The late positive potential was calculated as the mean amplitude in a 450 to 800 ms interval after feedback onset. Results show that subjects scoring higher on the “Concerns over mistakes” subscale displayed a smaller late positive potential to feedback for the incongruent condition. Furthermore, Tops et al. calculated the difference between the response times for correct trials following an error minus response times for correct trials following another correct response collapsed over congruency and examined the correlations between these difference scores, the “Concerns over mistakes” subscale, state arousal, and the late Pe. State arousal here was defined as the degree of self reported vigor minus fatigue (ranging from 0 to 14, higher values indicating more arousal).

Concern over mistakes negatively correlated with arousal ($r = -0.59$), while post-error slowing positively correlated with arousal ($r = 0.58$) and negatively with the late Pe for frontal electrodes ($r = -0.61$ for Fz, $r = -0.55$ for FCz). However, post-error slowing did not correlate significantly with “Concerns over mistakes” scores ($r = -0.31, p > 0.05$) or with the early Pe. In conclusion, Tops et al. note that they replicated the relation between the Pe and concern over mistakes, but at a different electrode compared to previous studies, possibly due to differences in feedback. Furthermore, they argue that their results support the notion that individuals scoring highly on the “Concerns over mistakes” subscale rely on disengagement coping to handle the shame and embarrassment of making mistakes.
Summary

The “Doubts about actions” subscale captures obsessional-like thinking in which people doubt their ability to accomplish tasks, expressing the extent to which a person is insecure about his or her performance and the degree to which someone will try to obtain optimal results. The correlation between this subscale with the ERN amplitude (Schrijvers et al., 2010), argued to reflect an action monitoring process enabling fast and flexible adjustment to changes in the environment, implies that these doubts drive more cognitive control and efforts to correct mistakes.

Intriguingly, the “Concerns over mistakes” subscale, representing the degree to which people experience distress over the occurrence of a mistake, does not appear to relate to action monitoring processes as indexed by the ERN (Schrijvers et al., 2010; Tops et al., 2013). However, the relation of the “Concerns over mistakes” subscale with the Pe matches nicely with this ERP’s hypothesized association with the emotional reaction following an error, though as noted by Overbeek et al. (2005) this association is not well replicated. A noteworthy aspect of these studies was the lack of any significant relationship between behavioral measures, in particular post-error slowing, and the measures of perfectionism.

Schrijvers et al. suggested that the lack of any relation may be due to how the flanker task is both perceptually and conceptually straightforward. They posit that the task may not place sufficient demands on processing resources such that complex control strategies or behavioral adjustments are needed for adequate performance. Other explanations include the small sample size and aggregate measures of post-error slowing, issues on which I will elaborate in later sections. As noted before, the ERN is believed to originate from the ACC (e.g. Debener et al., 2005; Dehaene et al.,
1994), whereas the Pe is believed to reflect activity from the anterior frontal gyrus and anterior insula (Ullsperger, Harsay, Wessel, & Ridderinkhof, 2010). Hence, the above studies suggest that perfectionistic traits may in part reflect differing activity in these brain regions, suggesting both the affective and cognitive aspects of the ACC along with the anterior frontal gyrus and anterior insula work to increase performance monitoring and negative affect, producing the greater concern and selective attention to errors seen in perfectionists. However, it is clear that more research is needed, as even simple factors such as the nature of feedback may greatly change the empirical results (Tops et al., 2013). In the following section, I discuss other behaviors and personality traits that may impact perfectionism.

1.2 Motivation and Personality

An important consideration yet to be addressed is how perfectionism is modulated by factors such as motivation.

1.2.1 Motivation and the ERN.

Though no one has directly examined the relation between perfectionism and motivation, a study by Pailing and Segalowitz (2004) examined how motivation impacted the ERN. Given the associations between the ERN and perfectionism, the results of this experiment can therefore provide important insights into perfectionism. Pailing and Segalowitz began by noting that the amplitude of the ERN increases when monetary incentives are offered for accuracy (e.g. Gehring et al., 1993). They argued that this suggests that the ERN includes an emotional or affective aspect. Therefore, the authors sought to further examine how ERN amplitude changes when different incentives for accuracy are offered. They also hypothesized that personality differences
would impact the ERN amplitude. Specifically, people less sensitive to motivational
states (i.e. people who are more conscientious) would exhibit a smaller impact on the
ERN from changes in incentives.

To test this, the authors had eighteen subjects complete a four choice letter task
with two response dimensions, one based on a vowel/consonant distinction, the other
based on an upper/lowercase distinction. Hence, subjects saw four categories of let-
ters (uppercase or lowercase, vowel or consonant). Each category was assigned to
either the left or right hand and the middle or index finger. Three types of errors
could therefore be made (the response involved the wrong hand, the wrong finger,
or both). There were four different incentive conditions. First, there was a no mo-
tivation condition, in which subjects were simply asked to perform the task quickly
and accurately. Then, there was an equal motivation condition, with equal incentives
offered for correct vowel and consonant identification and correct upper and lowercase
identification. Next, there were two unequal motivation conditions, where the incen-
tives favored one correct response over the other in a 3:1 ratio. Either the vowel and
consonant identification was favored over case, or upper and lowercase identification
was favored over letter type.

Each condition consisted of 512 trials with a maximum possible earning of 5
dollars. Both behavioral data and ERPs were recorded. Among several different
inventories administered, subjects also filled out a personality questionnaire consisting
of a hundred items from the International Personality Item Pool (Goldberg, 1999)
meant to measure the domains of neuroticism, extraversion, openness to experience,
conscientiousness, and agreeableness. Hence, the authors were able to examine the
association between personality characteristics, the type of incentive offered, and different measures of performances.

The incentives yielded significant differences between conditions for error rates as indicated by two repeated measures ANOVAs. Examining the no motivation and equal motivation conditions, there was a main effect where the number of errors was higher for the no motivation condition. Comparing the two unequal motivation conditions, there was a crossover interaction in which errors were higher for the response dimension less favored by the incentive ratio. The authors used a median split to create two groups of subjects, those scoring high or low on conscientiousness. A repeated measures ANOVA did not find a three way interaction (Condition x Error Type x Group) as predicted for the amplitude of the ERN. Pailing and Segalowitz then created a single measure reflecting the magnitude of the motivation effect for the ERN amplitude by summing difference scores for finger and hand errors (calculated by taking the difference between conditions favoring the relevant response dimension and discouraging it). They found a significant correlation \( r = -0.58, p < 0.05 \) between this new measure and conscientiousness scores. However, neuroticism scores accounted for more variance in this measure \( r = 0.74, p < 0.01 \).

Pailing and Segalowitz conclude that affective processes are reflected in the ERN, and it is moderated by personality. The authors suggest that contrary to their hypothesis, neuroticism may play a larger role in affective related changes in error rates, and hypothesize that locus of control may account for this finding. People high in conscientiousness tend to exhibit fewer external control perceptions, while in contrast people high in neuroticism tend to possess a greater number of external control beliefs.
Another possibility is that the anxiety involved in making a mistake played an important role in the differences observed in the ERN amplitude. This may explain why neuroticism accounted for more variance than conscientiousness. A viable alternative hypothesis is that individuals who exhibited more perfectionism had greater changes in ERN amplitude. This idea is supported by the fact that the “Concerns over mistakes” and “Doubts about actions” subscales for the FMPS have shown moderate sized correlations with neuroticism scores from the NEO-Five Factor Inventory, while the “Personal standards” subscale of the FMPS showed moderate correlations with conscientiousness (Parker & Stumpf, 1995).

The study by Pailing and Segalowitz (2004) highlights one of the challenges facing studies relating personality constructs to more fundamental cognitive processes. As emphasized by Shafran et al. (2002), when correlating behavioral and neural data with inventory responses, it can be difficult to determine whether the inventory accurately assesses a unidimensional trait or in fact measures several related traits. This is demonstrated by Pailing and Segalowitz’s finding that changes in ERN amplitude based on motivation was better accounted for by the neuroticism domain then the conscientiousness domain, contrary to their expectations. Moreover, several different plausible explanations could account for this finding. This highlights the importance of building a better understanding of the more fundamental cognitive processes involved, as this would allow better testing and elimination of competing hypotheses.

1.2.2 Negative Emotionality and the ERN.

Another study examining how the ERN is impacted by personality differences with implications for perfectionism was conducted by Luu, Collins, and Tucker (2000). The
authors tested the hypothesis that individuals scoring higher on measures of negative emotionality and affect would have a higher ERN amplitude, due to the distressing nature of errors and the association of negative consequences with incorrect responses.

The authors had 42 subjects complete a flanker task using letters (e.g. HHSHH). Again, individual deadlines were calculated, this time using 40 practice trials in which the initial deadline was set to 400 ms. The median response time from the last 30 blocks was then set as the new deadline for subsequent trials. Subjects started off with 3,200 points, and lost a point every 100 ms for a correct response that exceeded the deadline. An incorrect response (or no response at all) resulted in the loss of an additional eight points. At the end of the experiment, subjects received half a cent for every point they retained. Subjects were also informed at the beginning of the study that they would receive feedback about their performance relative to the other subjects (though at the end, everyone was told they had performed above average).

Subjects completed 800 trials broken up into blocks of 200, and behavioral data and ERPs were collected. Subjects also completed the Positive Affect Negative Affect Scale (PANAS; Watson, Clark, & Tellegen, 1988), and the Multidimensional Personality Questionnaire (MPQ; Tellegen, 1982). The PANAS contains 20 descriptions of affective states. An individual ranks, on a 5-point scale, how often he or she experienced one or more of the states in a given time period (which was the past few weeks for the present study). The MPQ is a 300 item personality questionnaire meant to measure 11 lower-order personality dimensions, whose scores can then be combined to derive scores for higher order dimensions like negative emotionality.

Two findings of Luu et al. (2000) are particularly relevant to my project. First, the authors split the subjects into two groups (those with high scores on negative
affect measures versus those with low) and examined how post-error slowing changed over time, looking at the mean response time of trials following an error for each of the four blocks of 200 trials. The authors found that both groups exhibited decreased post-error slowing as time increased. Second and most importantly, the authors found that those scoring highly on negative affect started off with greater post-error slowing compared to the low affect group, but then declined at a more rapid rate, thereby exhibiting less slowing by the end of the study compared to the low negative affect group. Furthermore, the authors found similar results with the ERN amplitude. Those with high negative affect exhibited greater amplitude in the ERN for the first set of 200 trials, but with the later blocks, individuals with high negative affect in fact had a lower average ERN amplitude compared to those with low negative affect.

Pailing and Segalowitz (2004) found similar results. They found that response times sped up across the four conditions for both their groups, indicating a practice effect. Also, the ERN amplitude tended to decrease from the first half of testing to the second half. These findings anticipate the hypothesis of Tops et al. (2013) regarding perfectionism: the possibility that subjects exhibiting higher concern with mistakes may regulate their emotions following errors, resulting in an attenuated ERN.

Furthermore, the results of Luu et al. may suggest one reason why examinations of post-error slowing from the studies on perfectionism found no significant relationships with the FMPS subscales. Studies typically aggregate response times in some form to create a measure of post-error slowing. For instances, Pieters et al. (2007) quantified post-error slowing as the difference between correct responses following incorrect responses and correct responses following correct responses. They then used a t-test to examine if the mean of this measure differed from zero. Such aggregation
may mask important patterns in post-error slowing over time, such as the possibility that subjects engage in more post-error slowing initially, and less by the end of the experiment. This in turn may prevent the detection of group differences. I discuss such implications in the next section and review in depth the usefulness of post-error behavior as a behavioral correlate.

1.3 Post-error Behavior

A major benefit of EEG methods is that they provide a more direct measure of neural processes, reflecting activity of large populations of neurons firing in tandem. As studies like Schrijvers et al. (2010) and Tops et al. (2013) demonstrate, this can allow identification of individual differences that would not necessarily be reflected in behavioral data. However, if behavioral measures in experiments like the flanker task could detect differing effects of perfectionism, this would be of great practical benefit to researchers, allowing insights into the functioning of perfectionism without having to resort to the expensive equipment needed for EEG methodology. Furthermore, the more information that can be gleaned from behavioral data, the better links that can be established between neural activity and overt behavior.

The data that appears most relevant to perfectionism in the context of forced-choice two-alternative tasks are error responses and the responses that follow them. The most commonly used measure in the literature, as seen in the Pieters et al. (2007), Schrijvers et al. (2010) and Tops et al. (2013) studies, is post-error slowing, which refers to how people exhibit slower response times in the trial (or trials) following an error (e.g. Laming, 1968; Rabbitt, 1966). Post-error slowing has been observed in a wide variety of tasks, from the aforementioned flanker task to the Stroop (e.g.
1.3.1 Early Work on Post-error Slowing

One of the earliest works on post-error slowing was by Rabbitt (1966), whose interest was piqued by how subjects commit errors in very simple and easy self-paced serial-response tasks, even when they are not stressed or fatigued. Rabbitt found that, in both a 4 choice and 10 choice task in which subjects had to correct any mistakes before being able to proceed to the next trial, errors (and their corrections) were significantly faster than correct responses. Most importantly, response times for the trial following an error were significantly slower than all other response types. At this early stage, Rabbitt posited that the post-error slowdown may be due to how...
the subject recognizes his or her mistake and slows down as a precaution to avoid further mistakes, or because the error interrupted a response rhythm the subject had established, resulting in slow performance until the rhythm could be established anew, but he noted that the data did not allow distinguishing between the hypotheses.

Laming (1968) also examined post-error behavior on three forced-choice two-alternative experiments. To examine various sequential effects, including the effect of an error on subsequent trials, Laming applied a multiple regression analysis to response times and choice proportions. First, he found that subjects were much less likely to repeat mistakes. Laming also found, like Rabbitt, that the response time on the next trial was significantly longer. Furthermore, Laming found that the size of the post-error slowing depended on the stimulus presented. If a different stimulus was presented compared to the one shown during error commission (i.e. the subject must make the same response as his or her erroneous response in the previous trial), the subject slows down even more.

Laming posited that errors were largely the result of irrelevant information sampled by the subject prior to the presentation of a signal during a state of temporal uncertainty. He suggested that the post-error slowdown represents how a subject delays his or her sampling of information to reduce the influence of irrelevant information and thereby improve accuracy, thereby leading to an increase in response times. Laming also suggested that the subject increases the threshold of information needed to make a response for the erroneous choice. This can account for why seeing the alternative stimulus in the next trial results in an even larger slowdown. Finally, Laming suggested that following correct responses, subjects slightly decrease their thresholds on the information needed to make responses each time (i.e. becoming
less cautious). It is interesting to note that Laming (1979), when re-examining the data, observed that such an account doesn’t fully explain the large slow-down, and pointed out that other accounts, such as a bias towards one response over the other, were also tenable.

Another oft-cited paper, with the memorable title “What does a man do after he makes an error? An analysis of response programming”, is a study by Rabbitt and Rodgers (1977). The authors had subjects complete 2 choice, 4 choice, and 8 choice tasks, varying the interval between a response and the next stimulus to be either 20 or 200 ms. One important finding was that the probability that subjects would make a mistake on the next trial following an error was significantly greater than chance, contrary to Rabbit (1966) and Laming’s (1968) results. Rabbitt and Rodgers, on closer examination, found that these double errors often occurred because subjects would make the correct response to the previous trial that they had first made an error, but this response was incorrect for the current trial. Furthermore, these “error correction responses” were much faster compared to all other response types. Only after removing these trials from post-error trials did Rabbitt and Rodgers consistently see post-error slowing.

Based on their results, Rabbitt and Rodgers conclude that the pooling of all post-error trials for analysis is ill-advised. Rabbitt and Rodgers also note that the original hypothesis that after any error subjects respond more slowly and cautiously to the next signal did not hold in this study, given the “error correction responses” that led to repeat mistakes.

Laming (1979) posits that the difference between the results in Laming’s (1968) experiments and Rabbitt and Rodgers’ (1977) work is due to the different intervals
between the response and stimulus between the studies. Laming’s (1968) experiments had intervals between 1.5 to 2 seconds, whereas Rabbitt and Rodgers had intervals of 20 or 200 ms. Laming (1979) also notes that an alternative interpretation instead of involuntary error corrections is that the short interval between responses and the next stimulus leads to an interaction between the memory trace for the stimulus presented in the previous trial and the new one being displayed, making it difficult for the subject both to discover that the error-trial response was wrong and to identify the new stimulus.

Several important observations can be drawn from this early work. First, these studies suggest that the degree of post-error slowing is sensitive to many external factors, such as the stimulus presented on the trial following an error and the interval between a response and stimulus. Furthermore, the studies demonstrate that there are different types of errors, and the type of error can greatly impact the latency of the response on the next trial, suggesting that pooling of error responses can mask important details. Finally, neither Rabbitt nor Laming commit to the theory that post-error slowing represents more cautious responding to avoid future errors. Despite this, as noted by Dutilh, Vandekerckhove, et al. (2012), authors continue to cite Rabbitt’s work in particular as an indication that post-error slowing represents a more cautious mode of responding. This holds true for the work examining perfectionism. For instance, Pieters et al. (2007) justify their use of post-error slowing by citing Rabbitt (1966), despite the later work of Rabbitt and Rodgers (1977).
1.3.2 Current Theories on Post-error Slowing

The early work on post-error behavior resulted in a mix of findings that often competed with each other, such as the findings of Rabbitt (1966) and Laming (1968) that the probability of a second error decreased after the first, versus Rabbitt and Rodgers’ (1977) discovery of an increased chance of making a second error. The current theories of post-error slowing also disagree on several points. Dutilh, Vandekerckhove, et al. (2012) cite a total of five different accounts for post-error slowing. For the researcher seeking to apply post-error slowing as a measure to aid in the study of individual differences, such as perfectionism for my project, this diversity of theories poses a distinct challenge for interpretation.

One of the most well-known and popular theories that provides an account for post-error slowing is the aforementioned conflict monitoring theory (Botvinick et al., 2001, 2004), which posits that specific neural structures, in particular the ACC, respond to conflict between incompatible response processes (for instance, the overriding of a prepotent response, selection among equally permissible responses, or an error response). This detection of conflict allows strategic adjustment of cognitive control to prevent conflict in subsequent performance. Post-error slowing is taken to reflect the result of this adjustment in cognitive control; following an error, subjects shift their performance strategy to a more conservative speed/accuracy balance (Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004). In particular, following detection of an error via the ACC and PMFC, a more cautious mode of responding is enacted by decreased activity in the response priming unit, leading to an increased motor threshold (Danielmeier & Ullsperger, 2011).
This account of a higher motor threshold corresponds closely to the early musings of Rabbitt (1966) and Laming (1968) suggesting that people speed up following a correct response until they make a mistake, and they subsequently become more cautious to avoid repeated mistakes. Several clinical studies and studies examining individual differences relying on post-error slowing interpret their results in the framework of the conflict monitoring theory (e.g. Hajcak, McDonald, & Simons, 2004; Paulus, Feinstein, Simmons, & Stein, 2004; Bogte, Flamma, van der Meere, & van Engeland, 2007; Sergeant & van der Meere, 1988). This also holds true for the studies by Pieters et al. (2007) and Schrijvers et al. (2010).

A more recent theory that is swiftly gaining popularity and competing with conflict theory’s account of post-error slowing instead posits that post-error slowing reflects surprise at unexpected events (e.g. Notebaert et al., 2009; Núñez Castellar, Kühn, Fias, & Notebaert, 2010; Wessel, Danielmeier, Morton, & Ullsperger, 2012). This line of reasoning arose in part because of conflict theory’s inability to explain certain findings. For instance, both Notebaert et al. (2009) and Núñez Castellar et al. (2010) note that conflict theory predicts both increasing response times and improved accuracy following errors. However, similar to Rabbitt and Rodger’s (1977) work, there have been several studies that have found reduced accuracy following errors (e.g. Hajcak, McDonald, & Simons, 2003; Hajcak & Simons, 2008).

Furthermore, pharmacological studies (Riba, Rodríguez-Fornells, Morte, Münte, & Barbanoj, 2005; Riba, Rodríguez-Fornells, Münte, & Barbanoj, 2005) have found that compounds that improve or reduce error detection do not impact the size of post-error slowing. Lesion studies in which there is damage to the brain structures believed to play an essential role in cognitive control, such as the lateral prefrontal
cortex and anterior cingulate cortex, have also been found to not affect post-error slowing (Gehring & Knight, 2000; Modirrousta & Fellows, 2008).

These findings have led researchers to posit that post-error slowing may instead reflect an orienting response to unexpected events. Notebaert et al. (2009) tested this by comparing performance across two conditions, one in which errors were infrequent versus the other in which correct responses were infrequent. The researchers observed post-error slowing with infrequent errors, but they instead observed post-correct slowing when errors were much more frequent. Núñez Castellar et al. (2010) found that post-error response times correlated with the P3 amplitude, which has been associated with the novelty processing of an orienting response (Friedman, Cycowicz, & Gaeta, 2001). Results like these led several researchers to posit that the brain circuitry involved in error processing is part of a system that generates predicted outcomes and adapts to unexpected events, be they errors or something else (J. W. Brown, 2013; Wessel et al., 2012).

Furthermore, though less-often cited and researched, the early theories on post-error slowing noted in section 1.3.1 remain viable explanations. First, the occurrence of an error may indeed negatively bias a subject against the choice made in error in following trials (Laming, 1968, 1979; Rabbitt & Rodgers, 1977). In other words, as pointed out by Dutilh, Vandekerckhove, et al. (2012) errors encourage response alterations and hinder response repetitions, impacting both response times and choice probabilities. It is also possible that following an error subjects more accurately controlled the onset in which they begin to sample information, reducing the variability in bias and therefore decreasing the chance that they may be biased towards an incorrect response (Laming, 1968, 1979). Finally, subjects may in fact delay processing of
evidence for the next trial following an error, perhaps in order to re-evaluate their performance or overcome disappointment (Rabbitt & Rodgers, 1977). Few researchers besides Dutilh, Vandekerckhove, et al. (2012) have examined these theories, and as such they are still feasible accounts of post-error slowing.

1.3.3 Support for Theories on Post-error Slowing

The two most dominant accounts for post-error slowing, suggesting that it occurs because of increased response caution or surprise, respectively, have both received mixed support. Supporting the increased response caution account, there have been several studies finding that post-error slowing is associated with increased accuracy (e.g. Rabbitt, 1966; Laming, 1968; Marco-Pallarés, Camara, Münte, & Rodríguez-Fornells, 2008; Danielmeier et al., 2011). Furthermore, several studies have shown that the degree of post-error slowing is predicted by the ERN and activity in the ACC and the prefrontal medial cortex (Danielmeier et al., 2011; Debener et al., 2005; Garavan, Ross, Murphy, Roche, & Stein, 2002; Gehring et al., 1993; Hester, Barre, Mattingley, Foxe, & Garavan, 2007) as well and increased motor inhibition (Danielmeier et al., 2011; King et al., 2010). However, in addition to the aforementioned studies finding issues with conflict theories account of post-error slowing, there have also been studies that have found not found any relationship between post-error slowing and prefrontal medial cortex activity (Gehring & Fencsik, 2001), while other studies have found that post-error slowing correlates with the Pe (Nieuwenhuis et al., 2001; Hajcak et al., 2003).

The orienting account, suggesting that post-error slowing occurs because of surprise at an infrequent event, can account for effects that the increased response caution
cannot, but has also received mixed support. For instance, in studies in which sub-
jects would occasionally receive false feedback indicating that they made a mistake,
it was found that they still only slowed for errors that they had truly committed
(De Bruijn, Hulstijn, Verkes, Ruijt, & Sabbe, 2004; Logan & Crump, 2010; Stein-
hauser & Kiesel, 2011). Finally, as mentioned earlier, the final three theories have
received little attention in current research, and therefore remain viable accounts with
a large degree of face validity, thereby warranting further testing. Hence, care must
be taken in interpreting post-error slowing, given the differing theoretical accounts.

1.3.4 Risks with Aggregation

Another major caveat in the use of post-error slowing as a behavioral marker of
different types of cognition lies in its quantification. Typically, quantification of post-
error slowing involves finding the difference between the mean response time for trials
following a correct response and trials following an error (Dutilh, van Ravenzwaaij,
et al., 2012). This value is often calculated per trial for each condition, which subse-
quently can be used in statistical tests like a t-test or ANOVA. Furthermore, certain
response types are often excluded, such as repeat mistakes (e.g. Rabbitt & Rodgers,
1977; Maier, Yeung, & Steinhauser, 2011). There are several troubling aspects to this
practice. First of all, several studies have shown that people commit several types of
errors in the course of a single task. Such an observation was made even in the earliest
work examining post-error behavior, as Rabbitt (1966, p. 264) states “errors made
during the same task may conceivably represent very different types of breakdown in
performance.”
To provide a plausible example, a person may have a mixture of two types of errors, ones due to distraction (slow errors, followed by fast corrections) and ones due to lack of caution (fast errors, followed by slow corrections). Aggregating over these will mask both the post-error slowing and speeding. Another possibility, noted by Dutilh, van Ravenzwaaij, et al. (2012), is that subjects may change their response strategies over time. A subject who starts out highly motivated but becomes fatigued by the end of a study will start out with fast and accurate responses, but by the end will have slower responses with more errors. Even if the subject does not engage in post-error slowing, because of the unbalanced occurrence of errors (during the subject’s slower, more fatigued stage of responding), the difference in mean response times for post-correct and post-error response times will erroneously indicate that post-error slowing has occurred. Hence, using aggregate measures of post-error slowing can lead to both spurious detection or lack of detection of slowing.

The issues regarding measures of post-error slowing may have all contributed to the lack of any relation found between perfectionism and post-error slowing, as neither the study by Schrijvers et al. (2010) nor Tops et al. (2013) found that the perfectionism subscales related to post-error slowing. However, Tops et al. (2013) appear to have used difference scores with mean response times, hence pooling error responses and aggregating, which as was mentioned is ill-advised. Schrijvers et al. (2010) do not specify the exact method in which post-error slowing was computed, merely stating that no effect using a repeated measures general linear model was found. Given the non-normality of response time data and the potential for sequential effects (Van Zandt, 2000a), this lack of details is problematic.
Furthermore, no study examining perfectionism and post-error behavior has considered other types of post-error adjustments besides slowing. Some studies examining general post-error behavior discuss post-error improvements in accuracy (e.g. Danielmeier & Ullsperger, 2011; Notebaert et al., 2009), but this doesn’t take into account the relationship between response times and accuracy. In turn, post-error speeding is less often discussed in the context of simple forced-choice two-alternative tasks like the flanker task.

In some ways, this lack of focus on post-error speeding is surprising, since it is easy to suggest interesting cognitive mechanisms underlying such a response. For example, a subject may be recruiting more top-down attentional modulations, allowing better processing of information and improved task performance. However, given that tasks like the flanker task are simple, relatively easy, and errors are typically infrequent, the nature of the task may be such that post-error slowing is much more likely, with only the occasional occurrence of post-error speeding. Furthermore, because errors in which subjects are distracted may produce much longer response times, such data may be removed during pre-analysis processing, thereby masking the effect. Nonetheless, consideration of multiple types of post-error adjustment may provide more information that could be potentially useful to evaluate individual differences in perfectionism.

At the very least, it is clear that the variability in post-error adjustments is due to a diverse set of sources, which typical methodology, such as taking the difference in means or carrying out an ANOVA, are inadequate to appropriately parse out. For measures of post-error behavior to be useful in aiding our understanding of perfectionism and behavior in general, such variability must be addressed. Fortunately, recent
work using mathematical models has shown great promise in quantifying post-error behavior in a theoretically relevant way, as well as using it to examine individual differences.

For instance, Dutilh, Vandekerckhove, et al. (2012) made use of the popular and successful diffusion model (Ratcliff, 1978; Ratcliff & McKoon, 2008) to test five different accounts for post-error slowing in a lexical decision task. The authors proposed that the different accounts map uniquely onto parameters in the diffusion model, and this one-to-one mapping between psychological processes and model parameters allows for an informative decomposition of post-error slowing to test the accounts. Specifically, the account of increased response caution would be reflected in higher thresholds on the amount of evidence needed for a choice following an error. The orienting account predicts that evidence accumulation towards the thresholds would be contaminated. The account that errors bias a person towards the other response would be reflected in a shift in the starting point of the evidence accumulation towards the other choice. Laming’s (1968, 1979) suggestion of decreased variability suggests that the trial to trial variability in the starting point of evidence accumulation would be decreased. Finally, Rabbitt and Rodger’s (1977) suggestion that subjects may have a delayed start-up in processing evidence for the next trial would be reflected in an increase in the non-decision components of the response time.

Dutilh, Vandekerckhove, et al. (2012) used data from 39 subjects who each contributed about 28,000 observations for a lexical decision task in which subjects indicated whether a letter string was a word or non-word. The large number of observations ensured sufficient error responses for fitting the model. The main finding was that support that subjects exhibited increased response caution following an error,
setting response thresholds to be higher. The authors also found that subjects following an error were biased towards indicating a letter string was a word, and during the overall experiment, subjects were biased towards the response they had made in the previous trial.

Dutilh, Forstmann, Vandekerckhove, and Wagenmakers (2013) used a simplified form of the diffusion model and a more robust quantification of post-error slowing (Dutilh, van Ravenzwaaij, et al., 2012) to examine individual differences specific to aging across different tasks. They had 15 undergraduates and 19 older subjects between 60 and 80 years old complete two sessions of a random moving dot task. Fits of the model showed that younger subjects exhibited longer non-decision times following an error, whereas older subjects also exhibited reduced evidence accumulation and higher response thresholds. Dutilh et al. also examined a combined dataset from two previous lexical decision tasks. Each of 102 younger and 58 older subjects contributed over 2100 observations. In the lexical decision task, both younger and older subjects exhibited a decline in evidence accumulation and increased response thresholds following an error. Older subjects also spent an increased amount of time on non-decision processes. Hence, the study by Dutilh and his collaborators emphasizes first, that the nature of post-error slowing and the cognitive processes involved differ across tasks and subjects, and second, that these differences can be determined via mathematical models of decision making. I will use a similar methodology to address questions about perfectionism, using models to examine post-error behavior across subjects with differing degrees of perfectionist characteristics. In the following section, I briefly review the key features of the most successful mathematical models.
of decision making, the further benefits of using such models, and the specific model I will apply to my data.

## 1.4 Quantitative Modeling

Most domains in psychology, including clinical psychology, rely on a verbal hypothesis-testing approach to theory development (Ratcliff, 1998). This verbal modeling approach involves testing a set of statements formulated from observations about behavior (Myung & Pitt, 2002). Verbal models can be used to identify key variables impacting behaviors of interest without speculating too far beyond what is observable in the data, which can be useful during the early stages of researching a new topic. However, there are diminishing returns with this approach as a lack of precision in the formulation of underlying mechanisms can make it difficult to generate predictions, discriminate between multiple interpretations, and assign meaningful magnitudes to effects. Moreover, the approach furnishes only a gross qualitative description of data, which can hide important details that can greatly impact interpretations. In contrast, quantitative models guided by theory can clearly indicate the assumptions being made on structural and functional characteristics of the phenomenon under study. Computational mechanisms can be specified, precise predictions can be made, and comparisons between falsifiable models can provide a stringent test of theories (Myung & Pitt, 2002).

Quantitative modeling has proven to be a powerful tool for research into decision making, especially in the realm of simple two-choice tasks. These tasks involve a fast (typically less than one second) decision between two response alternatives, and the measured behavioral variables are response times and choice. Such tasks
can be used to study a bevy of topics, including recognition memory (e.g. Ratcliff, 1978; Van Zandt, 2000b), lexical decisions (Ratcliff, Gomez, & McKoon, 2004), executive control (Verbruggen & Logan, 2009), and perceptual discrimination (Van Zandt, Colonius, & Proctor, 2000). A wide range of decision types can be examined, such as yes/no, old/new, same/different, categorization, two-alternative forced choice, and response signal (White, Ratcliff, Vasey, & McKoon, 2010c). Such tasks have seen a long history of application in clinical psychology (see White et al., 2010c for a review).

Sequential sampling models have had the greatest success in fitting the observed empirical patterns found in simple two choice tasks (e.g. Luce, 1986; Ratcliff & Smith, 2004; Van Zandt et al., 2000; Vickers, 1979). For these models it is assumed that the representation of stimuli in the central nervous system is inherently variable or noisy. To make a decision, people accumulate successive samples from the noisy stimulus representations in the course of a single trial until a threshold quantity of evidence is attained. The particular threshold that is reached determines which of the two alternatives is selected, and the amount of time taken to accumulate this evidence (i.e. the number of samples drawn, multiplied by the time it takes to accumulate a unit of evidence) determines the observed response time.

One of the main strengths of the sequential sampling models is their ability to describe performance on a task based on two features: the rate of accumulation, which describes the quality of information derived from the stimulus representations, and the decision thresholds, that describe response caution. The speed-accuracy trade-off can be accounted for by shifts in the decision thresholds. Thresholds that are set higher necessitate more information accumulation, producing more accurate
but slower responses, whereas thresholds set lower need less information to accumulate, producing faster but less accurate responses. Several models also incorporate a parameter representing the time spent on non-decision processes, such as the time required for the motor response to indicate a choice. In summary, as noted by Donkin, Brown, Heathcote, and Wagenmakers (2011) despite the large number of sequential sampling models almost all of them have these three common latent variables: the rate of evidence accumulation, response caution, and a non-decision component. Estimation of these variables can be used to make inferences about the causes of overt behavior in decision tasks in terms of latent psychological processes that are more meaningful than raw data summaries.

Studies of the cognitive mechanisms of perfectionism have had small sample sizes (ranging from 16 to 39 subjects). This small sample size may contribute to the difficulties in determining whether any associations between post-error slowing and perfectionism exist. However, power to detect group differences, which is important for these studies of perfectionism, is improved by using sequential sampling models with a large number of within-subject trails. A study by White, Ratcliff, Vasy, and McKoon (2010a) provides a good demonstration of how sequential sampling models can improve power and allow the detection of effects in a small sample size.

White et al. examined the empirical finding that individuals high in anxiety show enhanced processing of threatening information over nonthreatening information compared to people low in anxiety. Previous work had only been able to find reliably enhanced processing when there was competition for processing resources. White et al. (2010a) hypothesize that the reason that threat bias has so far only been observed in tasks with competing stimuli is due to threat bias possessing a small effect size
when there is no input competition. The authors argue that traditional analyses have not been sensitive enough to detect these effects, and point out several studies in which non-significant trends towards differences between low and high anxiety individuals have been found. The authors used a model-based approach to improve statistical power by applying the diffusion model (Ratcliff, 1978; Ratcliff & McKoon, 2008) which decomposes response time and choice data into separate underlying components representing response caution, bias, and quality of stimulus information as captured by separate model parameters (boundary separation, starting point, and drift rate, respectively).

The authors tested low and high anxiety individuals in a lexical decision task without input competition. The authors found that individuals high in anxiety had significantly higher drift rates for threatening compared to nontargeting words whereas individuals low in anxiety had lower drift rates. By contrast, while the response time and accuracy measures showed trends in the expected direction, these differences were not significant. Hence, it was only through the use of the diffusion model that the authors were able to detect the increased processing advantage for threatening words with high anxiety, allowing a more stringent test of the relevant cognitive theory as applied to a clinical question. Parallels can be drawn between the study by White et al. and the studies on the cognitive mechanisms of perfectionism. The latter studies were also trying to detect small effects between people with high and low levels of a clinical construct, yet so far no significant associations have been found for the behavioral measures relying on response time, specifically post-error slowing. However, sequential sampling models may provide the necessary boost in power to detect such effects, as was the case in the study by White et al. (2010a).
There are several current sequential sampling models under use, such as the diffusion model (Ratcliff, 1978; Ratcliff & McKoon, 2008), the leaky competing accumulator model (LCA; Usher & McClelland, 2001), the linear ballistic accumulator model (LBA; Brown & Heathcote, 2008), and the Poisson counter model (Townsend & Ashby, 1983; Van Zandt et al., 2000). Each of these models has its own advantages and disadvantages. Moreover, the models make very different assumptions regarding the decision process. For my project, I will focus on a race model which assumes a race between two independent diffusion processes, each of which has a single boundary.

1.4.1 The Diffusion Race Model

Race models, a specific subset of sequential sampling models, have had a long and often successful history in response time research (e.g. S. D. Brown & Heathcote, 2008; Townsend & Ashby, 1983; Usher & McClelland, 2001; Van Zandt et al., 2000; Vickers, 1979). The defining quality of a race model is that it posits that the decision process can be represented as a race between two or more stochastic accumulators towards separate decision thresholds. The accumulator that reaches its criterion or threshold first wins, determining the choice and response time. The specific assumptions made about the nature of the accumulators determines the predictions the model makes about choice probabilities and response time distributions. Further assumptions about the selective influence of different experimental manipulations can then be checked, providing insights into the cognitive architecture under study.

An attractive feature of a race model is its neural plausibility: the accumulators can be likened to the firing rates for separate neural populations processing evidence for each alternative. Indeed, several race models have successfully been applied to
fitting both behavioral data and single cell firing rates (e.g. Ratcliff, Cherian, & Segraves, 2003; Ratcliff, Hasegawa, Hasegawa, Smith, & Segraves, 2007; Ratcliff et al., 2011). For my project, the accumulators are described as Wiener diffusion processes (Logan, Van Zandt, Verbruggen, & Wagenmakers, 2014). The Wiener process is a popular choice for stochastic accumulation in sequential sampling models, as exemplified by the diffusion model (Ratcliff, 1978; Ratcliff & McKoon, 2008) and the LCA model (Usher & McClelland, 2001), and has proven to be very successful in accounting for response times and choice data across a wide variety of tasks.

Assume two accumulators, one for each alternative. The $i$\textsuperscript{th} accumulator is a Wiener diffusion process with a drift rate $\xi_i > 0$, a threshold (absorbing boundary) $\kappa_i > 0$, and a starting point fixed at 0. The within-trial variability, or drift coefficient $\sigma$, is fixed at 1 for identification purposes. Therefore, the finishing time $T_i$ for the $i$\textsuperscript{th} accumulator follows a Wald or inverse-Gaussian distribution. For an observed finishing time $T_i = t$, the cumulative distribution function of the Wald is

\[
F(t|\kappa_i, \xi_i) = \Phi \left( \frac{\kappa_i}{\sqrt{t}} \left[ \frac{\xi_i t}{\kappa_i} - 1 \right] \right) + \exp \left\{ 2\xi_i \kappa_i \right\} \Phi \left( -\frac{\kappa_i}{\sqrt{t}} \left[ \frac{\xi_i t}{\kappa_i} + 1 \right] \right),
\]

(1.1)

where $\Phi$ is the cumulative distribution function for the standard normal distribution, and we take it as implicit that the support of $F$ is the positive real line. In turn, the probability density function is

\[
f(t|\kappa_i, \xi_i) = \frac{\kappa_i}{\sqrt{2\pi t^3}} \exp \left\{ -\frac{1}{2t} \left( \frac{\xi_i t}{\kappa_i} - \kappa_i \right)^2 \right\}.
\]

(1.2)

Because the observed response time and choice is determined by the accumulator that wins the race, the distribution of the finishing time that wins the race and determines the response is the distribution of the minimum of the two Wald finishing times. Hence, the joint likelihood of observing a response time $T = t$ given that the
Figure 1.1: The accumulation process during a single trial for the diffusion race model given drift rates \( \xi_i \) and \( \xi_j \), thresholds \( \kappa_i \) and \( \kappa_j \) (set equal to each other for presentational clarity), and a non-decision component \( \tau \) (a shift added to the decision time). Here, the accumulator for choice \( i \), which has the higher drift rate, reaches its threshold first, producing the observed response time and subsequent choice of \( i \).
accumulator for choice $i$ finished before the accumulator for choice $j$ is

$$g(t|\kappa_i, \xi_i, \kappa_j, \xi_j) = f(t|\kappa_i, \xi_i) [1 - F(t|\kappa_j, \xi_j)]. \quad (1.3)$$

Note that equation 1.3 indicates that the choice $i$ “won” (i.e. the corresponding accumulator reached its threshold first) by the ordering of the parameters in the conditioning.

Figure 1.1 provides a visual example of the diffusion race model. This visualization makes it easy to see how the stochastic accumulators for choices $i$ and $j$ race each other with rates $\xi_i$ and $\xi_j$ to reach their respective thresholds $\kappa_i$ and $\kappa_j$, set here to be equal. In this case, the accumulator for choice $i$ reaches its threshold first, winning the race and determining the choice and decision time. The observed response time is then a sum of the decision time and a non-decision time $\tau$.

As I mentioned earlier, a major strength of sequential sampling models is that the parameters have specific psychological interpretations that can be selectively influenced by different experimental manipulations. In equation 1.3, the drift rate $\xi_i$ can be interpreted as representing the quality of information driving evidence accumulation on accumulator $i$ when the stimulus for choice $i$ is actually shown. A more easily-discriminated stimulus associated with $i$ will result in a higher value for $\xi_i$ (and a lower value for $\xi_j$). This parameter can also capture task engagement. A subject who is paying more attention will better sample information from the stimulus associated with choice $i$, again producing a higher value of $\xi_i$.

By contrast, the parameters $\kappa_i$ and $\kappa_j$ are assumed to be under a subject’s control, and variations in $\kappa_i$ and $\kappa_j$ capture differences in response strategy. For example, higher values of $\kappa_i$ and $\kappa_j$ indicates a more cautious mode of responding, whereas lower values indicate a greater emphasis on speed. Biased responding is represented
by $\kappa_i \neq \kappa_j$, which increases the likelihood that the subject picks the choice associated with the accumulator with the lower threshold. To incorporate the time taken for the non-decision processes, as is the case with other sequential sampling models, the shift parameter $\tau$ can be subtracted from the finishing time; giving a joint probability density function

$$g(t|\kappa_i, \xi_i, \kappa_j, \xi_j, \tau) = f(t - \tau|\kappa_i, \xi_i) \left[1 - F(t - \tau|\kappa_j, \xi_j)\right]. \quad (1.4)$$

Equation 1.4 is the basis for both maximum likelihood estimation and Bayesian estimation of the model parameters $\kappa_i$, $\kappa_j$, $\xi_i$, $\xi_j$, and $\tau$. Estimates of the model parameters in turn allow insights in the cognitive processes involved in the decision process.

The diffusion race model was recently successfully applied to address questions about stop-signal tasks in a paper by Logan et al. (2014). 6 subjects each completed a 2-choice, 4-choice, and 6-choice task in which they had to indicate which word was present on screen out of a set of 2, 4, or 6 (respective to the task). In a quarter of the trials, subjects would receive an auditory stop signal indicating that they should not respond to the stimulus. The stop signal was played following a delay after stimulus presentation that ranged from 35 percent to 95 percent of the subject’s mean response time. The authors found that the model was able to account for subjects’ accuracy, the response time distributions for correct choices, and the stop-signal response time, which cannot be directly observed. However, the model predicted slower error responses compared to empirical data, which Logan et al. (2014) attribute to the low error rate. Simulations indicate that with parameter variability in thresholds and a high error rate, the model can predict fast errors.
There are other race models that I could use that still assume the stochastic process for the accumulators is described by a diffusion process. For example, Ratcliff et al. (2007) and Ratcliff et al. (2011) used a race model with racing Ornstein-Uhlenbeck processes, a diffusion process in which evidence accumulation can decay with time. This model demonstrated good fit to both behavioral and neural data from primates. The LCA model (Usher & McClelland, 2001) also posits racing diffusion processes with both a decay term and an inhibition term, where the activity of one accumulator can suppress activity of the other accumulator. The LCA model has neural plausibility and has successfully fit data from several experiments; for instance a LCA without leakage has been applied to data from a stop-signal task (Boucher, Palmeni, Logan, & Schall, 2007). While these models are more neurally plausible and flexible, they possess no closed-form solutions for the equations used in fitting the model, requiring computationally intensive and less stable simulations to estimate model parameters. The diffusion race model is an attractive alternative in that it is easier to fit while still providing adequate fit to data in a neurally plausible framework, allowing me to rigorously test my hypotheses on perfectionism and post-error behavior.

1.5 Summary

Perfectionism is an ill-defined construct that nonetheless plays a major role in psychopathology. Our understanding and definition of this construct can benefit from examining its relation to specific behaviors like decision making, and better explicating the underlying cognitive processes involved in perfectionistic behavior. Therefore
for my project I apply the power of sequential sampling models to explore individual differences in perfectionism in the context of simple choice, examining whether post-error slowing in a Simon task can reveal important insights about perfectionism.
In the following, I describe the experimental design used to examine my hypotheses. Because the goal of my project is to examine perfectionism and its influence on post-error behavior, I needed a task well-known to produce post-error adjustments. While previous studies studying perfectionism have relied on the Eriksen flanker task, I instead used a variant of the Simon task. Task difficulty in previous studies was manipulated by using an individualized response time deadline to produce a higher error rate (Pieters et al., 2007; Schrijvers et al., 2010; Tops et al., 2013). I chose not to do so, as such a procedure would require a host of additional modeling assumptions. Such deadlines can be thought of as imposing a speed emphasis on subjects, which impacts the response thresholds in sequential sampling models. However, because the deadline changes every block and is contingent on a subject’s previous performance, this introduces a dependency structure beyond the scope of this project. In contrast, the Simon task allowed me greater control over the nature of the stimuli under display, in turn allowing greater control over task difficulty without resorting to an individual response time deadline. Performance on the Simon task still exhibits interference effects similar to the Eriksen flanker task, and the Simon task has been found to produce post-error slowing (King et al., 2010; Danielmeier et al., 2011), making it an appropriate task to address my research questions.
2.1 Subjects

Data was collected from two different groups of subjects. First, a pilot study was conducted using five graduate students from the Quantitative Psychology program at Ohio State University (OSU), four of whom completed two sessions of the study, while the fifth completed a single session. Then the main study was conducted using twenty-nine undergraduates recruited from the introductory psychology courses through the Research Experience Program (REP) at OSU. The subjects were 18 years or older, and possessed normal or corrected-to-normal vision. Seven subjects completed a single session of the experiment, while twenty-two subjects completed an additional session following the first.

2.2 Materials

Subjects completed the “Concerns over mistakes”, “Personal standards”, and “Doubts about actions” subscales from the FMPS (Frost et al., 1990). Frost et al. report that the Cronbach’s alpha for these three subscales is 0.88, 0.83, and 0.77 respectively, based on a sample of 178 women from an introductory psychology course. The correlation between the “Concerns over mistakes” and “Personal standards” subscales was $r = 0.47$, while the correlation between the “Concerns over mistakes” and “Doubts about actions” subscales was $r = 0.47$, and finally the correlation between the “Personal standards” and “Doubts about actions” subscales was $r = 0.24$.

Subjects also completed the A-State form of the STAI (Spielberger et al., 1970). The scale is designed to measure state anxiety, the transitory feeling of tension and apprehension that can fluctuate over time and vary in intensity (Vigneau & Cormier, 2008). The form consists of twenty statements asking subjects how they feel at the
current moment of time. Subjects read items like ‘I feel calm’ or ‘I am jittery’ and then pick the response category representing how they feel using the four point scale: ‘not at all’, ‘somewhat’, ‘moderately so’, or ‘very much so’. Spielberger et al. examined the reliability of the STAI across three samples of students; Cronbach’s alpha for the STAI A-State form ranged from 0.83 to 0.92 across these three samples. Test-retest reliability of the form, administered to a sample of undergraduates, was low, ranging from 0.16 to 0.54 across durations of one hour, day, and one hundred and four days. As noted by Spielberger and his collaborators, the low stability of the A-State form is expected for a measure meant to be influenced by situational factors.

There were six questions at the very end of a session, following the main task and the two inventories. The subjects were reminded that their data were anonymous, and their honest responses would greatly help. Subjects were first asked whether they were right-handed, left-handed, or ambidextrous (to help check for response biases based on handedness). Next, they were asked to indicate how many hours of sleep they got last night (as less sleep may contribute to a higher error rate). Subjects then indicated how difficult they found the task initially and by the end using the following scale: ‘Easy’, ‘balanced’, or ‘hard.’ subjects were also asked how well they were able to stay focused during the task, using the response categories: ‘Rarely’, ‘half of the time’, ‘most of the time’, and ‘all of the time.’ Finally, subjects were asked to rate their experience during the study, using the scale: ‘Very bad’, ‘bad’, ‘Neither bad or good’, ‘good’, and ‘very good.’
Instructions:
Press the "a" key when the rectangle is:
Be as accurate and quick as possible!

Figure 2.1: The instructions, displayed stimuli, and timing using two example trials of the start of an experimental block.
2.3 Design

Subjects in my study completed a variant of the Simon task (Simon & Wolf, 1963), in which a small square (10 by 10 pixels) flashed either on the left or right hand side of the monitor. Subjects were tasked with indicating whether the square was red or orange, ignoring the side in which the stimulus was displayed. These colors were chosen as they are highly similar and more confusable, thereby increasing task difficulty and promoting the occurrence of errors. In half of the blocks, subjects needed to press the “a” key when the square was orange and the “l” key when it was red, whereas in the remaining blocks the key assignments were reversed.

Incongruent trials occurred infrequently. Hence, of the one hundred trials in a block, thirty-five orange squares and thirty-five red squares were displayed on the side congruent with the key assigned to each color. In turn, fifteen orange squares and fifteen red squares were displayed on the side incongruent with their respective assigned key. For instance, if a subject was instructed to press the “a” key when the square was orange, a congruent trial would refer to when an orange square was displayed on the left, whereas an incongruent trial would refer to when an orange square was displayed on the right. Furthermore, a square’s distance from the center on the x-axis varied from trial to trial according to a normal distribution. On average, squares would appear one hundred pixels off center with a standard deviation of twenty pixels. This variability was intended to induce further difficulty, because subjects would not be able to anticipate the exact location that the square would appear on either side.

The $n^{th}$ trial started with the trial number appearing for 250 ms in the center of the monitor, informing subjects of their progress and also functioning as a fixation
point. Following a 200 ms gap, the target stimulus flashed for 25 ms. Subjects then needed to make a response using the “a” and “l” keys. Note that a response made before and during the presentation of the stimuli was not registered. Following the response, there was a 50 ms gap, and then the next trial began. As the key assignments changed from block to block, and there was no trial-by-trial feedback on performance, a potential risk was that a subject would forget the response mappings. To avoid this, at the bottom center of the monitor a small rectangle was always displayed during a block, indicating the color currently assigned to the “a” key. The distance of this rectangle from the center of the monitor, and the fact that it was always present (even during the interstimulus gaps) prevented any interference effects. Figure 2.1 provides examples of the instructions shown at the start and two example trials.

In half of the blocks, subjects earned or lost points based on their performance. For incorrect responses subjects lost one point, while for correct responses they gained a quarter of a point. At the end of each block of trials, subjects were informed of their percentage accuracy, and, when appropriate, their current number of points earned. The proportion of blocks they completed were also shown. Subjects earned money for their participation, based on their performance on the task. Subjects received six dollars plus an additional four cents for every point made.
2.4 Procedure

The experiment was conducted using the program Opensesame, version 0.27 (Mathôt, Schreij, & Theeuwes, 2012). Stimuli were presented on a standard CRT desktop monitor, and subjects’ responses were recorded using a standard QWERTY keyboard. All keytops but those for the numbers, backspace, enter, “a”, and “l” keys were removed.

Subjects were tested individually in a quiet room, and completed their experimental sessions within a one week period. Upon arrival they were told that the study was confidential and voluntary, such that they may have left at any time with no consequences and they would still receive compensation for their participation. The instructions for the experiment were displayed on the monitor in addition to being read out loud by the experimenter.

For each session, subjects first completed two blocks each consisting of 20 congruent and 10 incongruent trials. In the first block, the “a” key was assigned to orange squares, while in the second block it was assigned to red squares. Subjects were then asked to give a number between 0 and 100 percent as an estimate of their accuracy for the rest of the task. Additionally, subjects indicated how well they would like to perform relative to others with the following scale: “Doesn’t matter”, “average is fine”, “above average”, or “greatly above average.” Finally, subjects then indicated how well they thought they would actually perform relative to others, based on the following response categories: “Greatly below average “, “below average,” “average,” “above average,” or “greatly above average.”

Subsequently, subjects completed sixteen blocks of one hundred trials each. In eight of the blocks they earned points (the motivation condition), while in the remaining eight they earned no points (the neutral condition). For each set of eight
blocks, in four of the blocks subjects were reminded that they should press the “a” key when the square was orange, while in the remaining blocks they were told to press the “a” key when the square was red. The color corresponding to the “l” key was not explicitly mentioned to help subjects avoid potentially confusing the two responses. The order of the blocks was randomized.

Following the experimental blocks, subjects completed the FMPS subscales and the STAI-A form. Finally, they completed the final set of six questions inquiring about whether they were right-handed, how much sleep they got the preceding night, how hard they found the tasks at the beginning and end, how focused they were during the task, and their overall experience in the study. A single experimental session took about 45 minutes.

Following the data collection, a mixture model was developed in order to account for the patterns of behavior observed in the pilot study data, which was then fit to the data from the main study in order to examine whether perfectionist characteristics could predict changes in parameter values across subjects. In the next chapter I detail the development of the model and its structure.
Chapter 3: The Model

Decision making is heavily influenced by personality and individual differences. Perfectionism stands out given its major role in psychopathologies such as depression and eating disorders and its contribution to noted deficits in decision making performance (Pieters et al., 2007; Schrijvers et al., 2010). Past methodology that relied on aggregate measures of post-error slowing is not optimal for uncovering why these deficits arise. Fortunately, quantitative modeling of the processes involved in decision making can provide insights into how decision processes unfold in both normal and psychopathological populations and can provide additional insights into the nature of mental illness. I therefore use the diffusion race model to provide insight in the role of perfectionist characteristics in post-error behavior for a forced-choice two-alternative task.

The specific problems addressed in this thesis are:

- defining a model that can account for the interference effect seen in the Simon task;

- controlling for factors like motivation and global changes in performance to avoid confounds when measuring post-error behavior;
• identifying whether a subject engages in several different types of post-error
  adjustments within a single experimental session; and

• determining whether perfectionist characteristics predict greater post-error slow-
  ing.

To address these problems, I used a mixture model combining two racing diffusion
processes and a log-normal process that explicitly accounts for changes in performance
across blocks of trials due to stimulus type, task demands, and several types of errors.
In the following, I discuss the modeling challenges I faced, the different models I ex-
plored, and the structure of the final model I selected to fit to my data. In developing
the final model, because the diffusion race model has never been fit to the Simon task
before, I made extensive use of the data collected in my pilot study to gain insight
into which mechanisms were necessary to include. To examine post-error behavior,
I tested several variants of the model with different covariates. I then used point
estimates of the model parameters to examine associations with perfectionism.

3.1 Fitting the Simon task

To account for post-error behavior, it is first necessary to fit typical response time
distributions produced by the Simon task. In the Simon task, different patterns of
data are expected depending on the stimulus attributes to which subjects attend and
what types of trials are intermixed in a given block (Proctor, Yamaguchi, Dutt, &
Gonzalez, 2013). For my experimental design, subjects were instructed to ignore the
stimulus location and attend only to the color of the stimulus, and only two types
of trials were present: those in which response location was congruent with stimulus
location, and those in which it was not. It is predicted that the subjects performing
my task will exhibit a strong Simon effect by making faster correct responses on average to congruent stimuli, but faster error responses on average to incongruent stimuli. The Simon effect influences the distribution of response times conditioned on accuracy and congruency (e.g., Proctor et al., 2013; Servant, Montagnini, & Burle, 2014). In the incongruent condition, the distribution of error response times is shifted downward relative to the distribution of correct response times. By contrast, the response time distribution for errors in the congruent condition is shifted upward relative to the distribution of correct response times. These distributional shifts pose a particularly thorny problem for sequential sampling models (Servant et al., 2014), as evidenced by my initial attempts at fitting the pilot study data.

### 3.1.1 Modeling Challenges

My initial model posited that the Simon effect could be accounted for by a single decision process with shifts in the drift rate and thresholds based on congruency. Figure 3.1 provides a visual representation of the decision process when a red square was displayed under the two different congruency conditions. The average path of the evidence accumulation for each choice is shown as an arrow. The slopes of these lines are determined by the drift rates for each accumulator, and the decision time is determined by how long it takes an accumulator to reach its respective threshold. Dashed lines denote the path and threshold for the accumulator corresponding to an incongruent choice. The subscripts $C$ and $I$ denote the congruent and incongruent condition, respectively, while the subscripts $R$ and $W$ indicate the right and wrong choices. In the diagram, because a red square was shown, the right choice is “red”, and the wrong choice is “orange”. Therefore, the rate of evidence accumulation for
Figure 3.1: The decision process in both the congruent (left-hand side) and incongruent (right-hand side) condition as proposed in my original model. The arrows denote the average path of the evidence accumulation, while the color of the arrow indicates to which choice the accumulator corresponds. The example assumes a red stimulus was shown.
making a “red” choice is determined by the drift rates $\xi_{CR}$ and $\xi_{IR}$, which for a well-performing subject at least, will both be higher relative to the rates of evidence accumulation for picking “orange” across congruency conditions. This is represented in the diagram by the higher slopes for the “red” accumulator. However, in the incongruent condition, due to interference from the spatial location I expected $\xi_{IR} < \xi_{CR}$ and $x_{IW} > x_{CW}$. Furthermore, subjects would be biased towards a congruent response, irrespective of the color of the displayed stimulus. Therefore in the diagram there are two thresholds $\kappa_C$ and $\kappa_I$. When the red stimulus was congruent with response location, the accumulator for picking “red” terminated at $\kappa_C$, but when the red stimulus was incongruent, the accumulator instead terminated at $\kappa_I$.

This model had a major limitation in that it was unclear how subjects could guess whether an upcoming trial would be congruent or incongruent, and adjust their response thresholds accordingly. Simulations and fits to the pilot study data revealed another major limitation: the model was unable to predict fast errors (by fast errors, I mean an error response time distribution shifted downward relative to the distribution of correct response times). The inability to predict fast errors is demonstrated in figure 3.2. The figure is a quantile-probability plot, which is a useful way to compare model predictions regarding response time distributions and accuracy relative to observed results (e.g., Ratcliff & McKoon, 2008). The 0.1, 0.3, 0.5, 0.7, and 0.9 quantiles for the response times in a given condition and response type (correct versus error) are plotted against the probability of the response (where a probability greater than 0.5 indicates correct responses). The ellipses overlaid on the plot represent the 95% credible intervals for the model predictions of the quantiles and corresponding response probabilities (details on how these intervals are calculated...
Figure 3.2: Quantile-probability plot showing the 3rd pilot subject's observed 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles on the y-axis and probability of a correct or incorrect response on the x-axis for the congruent (circles) and incongruent (crosses) conditions. The 95% credible intervals for the model predictions are shown as ellipses.
are given in Chapter 4). The green ellipses are the intervals for the predicted median response time. Critically, the plot reveals that the original model cannot predict distributions for error response times that are shifted below those of the correct responses. Hence, while the model correctly predicts response probabilities, it fails in fitting adequately the response times.

The challenges of fitting both fast and slow errors has a long history in the sequential sampling model literature (e.g., Luce, 1986; Ratcliff & Smith, 2004). Most successful models address this issue through the inclusion of parameter variability. For instance, Ratcliff and Rouder (1998) demonstrated that by allowing both the drift rate and the starting point of the diffusion process to vary from trial to trial, the model was able to account for conditions in which mean error times were faster than correct times as well as conditions in which mean error times were slower than to correct times. The LBA model (Brown & Heathcote, 2008) also handles faster and slower mean error times by positing trial-to-trial variability in drift rates and starting points (in fact, this is the only type of variability assumed in the model). Variability in starting points allows these models to predict faster average errors, while the variability in drift rates allows them to predict slower average errors.

The diffusion race model has a closed-form solution for variability in thresholds (Logan et al., 2014), which is equivalent to letting the starting points vary from trial to trial. In this formulation, the threshold varies from trial to trial according to a uniform distribution with a mean equal to $\kappa$ and a range $\alpha$, such that the threshold for a given trial may lie anywhere between $\kappa \pm \alpha$. However, even when parameter variability in thresholds is included, the model still has trouble predicting the degree to which the error response time distribution is shifted downwards relative
to the correct distribution. Moreover, the model’s efforts to capture this shift result in collateral distortion of the variance of the error response time distribution. When fitting the LBA and diffusion race models with starting point and threshold variability to the pilot study data, the models were only able to handle these shifts by way of extremely high degrees of parameter variability. This produced predicted response time distributions with abnormal characteristics.

Figure 3.3 shows the joint correct and error response time distributions for the diffusion race model when threshold variability was high (i.e. $\alpha \approx \kappa$). While the predicted distributions exhibited the relative shift between error and correct response times, the predicted density for error response times is almost triangular, with a rapid rise in the left tail and a much slower drop-off in the right tail. This does not match typical empirical densities for response time distributions. Another major limitation of parameter variability in thresholds (or starting points), is that it would predict fast errors in both the congruent and incongruent conditions. However, fast errors are not typically observed in the congruent condition for the Simon task (e.g. Servant et al., 2014). For these reasons, I did not incorporate parameter variability into my modeling assumptions to account for the Simon effect. The model that fit my data best was a mixture model consisting of two different racing diffusion processes and a log-normal process reflecting slow responses. In the following sections, I review the diffusion race model, discuss why I incorporated a log-normal component, and then discuss the additional covariates and parameters I used to fit my data.
Figure 3.3: The predicted correct (solid lines) and incorrect (dashed lines) joint response time densities based on the racing diffusion model given extremely high threshold variability for errors.
3.1.2 Review of the Diffusion Race Model

First, recall that in the diffusion race model each accumulator is a Wiener diffusion process with drift rate $\xi > 0$, threshold $\kappa > 0$, drift coefficient $\sigma = 1$, and a starting point fixed at 0. The finishing time of each accumulator follows a Wald distribution, making the distribution of the finishing time that wins the race the distribution of the minimum of the two Wald finishing times. Let $t^j = \{t^j_1, t^j_2, \ldots, t^j_{N_j}\}$ be a vector of response times for Subject $j$, with $t^j_i \in (0, \infty)$ and $N_j$ equal to the number of observations provided by subject $j$. In turn, let $y^j = \{y^j_1, y^j_2, \ldots, y^j_{N_j}\}$ be a vector of choices corresponding to $t^j$ with $y^j_i \in \{0, 1\}$, where 1 indicates choosing choice A (the color red in this case) and 0 indicates choosing choice B (the color orange). Let $\kappa^j_A$ and $\kappa^j_B$ be vectors of parameters relating to response thresholds and $\xi^j_A$ and $\xi^j_B$ be vectors of parameters relating to drift rates for the “red” and “orange” response accumulators, respectively. The parameters $\kappa^j_A$, $\kappa^j_B$, $\xi^j_A$, and $\xi^j_B$ have support over $(0, \infty)$. Finally, let $\tau^j$ be the residual latency for Subject $j$. Here $\tau^j$ has support over $[0, \min(t^j))$. I then specify the joint density of the data for the diffusion race model as

$$(t^j, y^j) \sim f_D(t^j, y^j | \kappa^j_A, \xi^j_A, \kappa^j_B, \xi^j_B, \tau^j), \quad (3.1)$$

where, from Equation 1.4,

$$f_D(t^j, y^j | \kappa^j_A, \xi^j_A, \kappa^j_B, \xi^j_B, \tau^j) = \prod_{y^j_i = A} g(t^j_i | \kappa^j_A, \xi^j_A, \kappa^j_B, \xi^j_B, \tau^j) \times \prod_{y^j_k = B} g(t^j_k | \kappa^j_B, \xi^j_B, \kappa^j_A, \xi^j_A, \tau^j). \quad (3.2)$$

For clarity, in the rest of the discussion I will focus on a single subject and drop the superscript $j$. Furthermore, for conciseness and generality, I will rely on matrix
algebra notation to represent the parameters of the model in terms of the covari- 
ates representing experimental conditions and individual differences. Specifically, I 
decompose the parameters $\kappa_A$, $\kappa_B$, $\xi_A$, and $\xi_B$ into the weighted sums of covariates 
in a $N \times P$ design matrix $X$ and a column vector $\beta$ (where $P$ represents the num- 
ber of coefficients). In the following I subscript the design and coefficient matrices 
appropriately to clearly indicate when I am referring to the thresholds or drift rates 
of a given racing diffusion process. I now describe the decision process that drives a 
subject’s choice based on the color of a stimulus.

### 3.1.3 The Decision Process

For the decision process, I assumed that a person’s drift rates were determined by 
which choice was correct (i.e. choosing the correct color) and whether the condition 
was congruent. I specified four drift rates: $\xi_{CR}$, $\xi_{CW}$, $\xi_{IR}$, and $\xi_{IW}$. As before, the 
subscripts $C$ and $I$ refer to the congruent and incongruent conditions respectively, 
while the subscripts $R$ and $W$ refer to the correct and incorrect choice, respectively. 
As an example, $\xi_{CR}$ determines the rate of the accumulation process for the correct 
choice in the congruent condition. Note this formulation means that the drift rate for 
a red stimulus and an orange stimulus are equivalent, corresponding to an assumption 
that the two colors are equally easy to process. While this is not necessarily true, 
differences in processing of colors is not the focus of this project, and early fitting 
efforts found that differing drift rates based on color overlapped to a large degree. 
Therefore, this assumption was deemed reasonable.

Let $X_{\xi_{AD}}$ and $X_{\xi_{BD}}$ be the design matrices for the parameters $\xi_A$ and $\xi_B$ respec- 
tively, the subscript $D$ indicating that this refers to the decision process. Let $\beta_{\xi_{AD}}$
and $\beta_{\xi_{BD}}$ by column vectors of coefficients. Furthermore, to describe the dummy coding I employed for the design matrices, let the set of observations $\{y_1, y_2, y_3, y_4\}$ in this case refer to a pair of observations from the congruent condition followed by a pair of observations from the incongruent condition. Moreover, for $y_1$ and $y_3$, the displayed stimulus was red, whereas for $y_2$ and $y_4$, the displayed stimulus was orange. Take observation $y_1$ as an example. Because this observation is from a congruent trial in which the displayed stimulus was red, the drift rate for the “red” response accumulator $\xi_A$ should equal $\xi_{CR}$. In turn, the drift rate for the “orange” response accumulator $\xi_B$ should equal $\xi_{CW}$. In contrast, observation $y_4$ is from an incongruent trial in which an orange stimulus was shown. Here $\xi_A$ should equal $\xi_{IW}$, while $\xi_B$ should equal $\xi_{IR}$. These circumstances can be achieved by the following dummy coding and ordering of coefficients (where the subscripts $A$ and $B$ indicate a red or orange stimulus was shown respectively):

$$\xi_A = X_{\xi_{AD}} \beta_{\xi_{AD}} = \begin{bmatrix} C_A & C_B & I_A & I_B \\ y_1 & 1 & 0 & 0 \\ y_2 & 0 & 1 & 0 \\ y_3 & 0 & 0 & 1 \\ y_4 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \xi_{CR} \\ \xi_{CW} \\ \xi_{IR} \\ \xi_{IW} \end{bmatrix},$$

and

$$\xi_B = X_{\xi_{BD}} \beta_{\xi_{BD}} = \begin{bmatrix} C_A & C_B & I_A & I_B \\ y_1 & 1 & 0 & 0 \\ y_2 & 0 & 1 & 0 \\ y_3 & 0 & 0 & 1 \\ y_4 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \xi_{CW} \\ \xi_{CR} \\ \xi_{IW} \\ \xi_{IR} \end{bmatrix}. $$

Note that in later sections, I will add additional columns to the design matrices and add further coefficients.

I performed similar decompositions of the parameters $\kappa_A$ and $\kappa_B$, and set $\kappa_A = \kappa_B$. This constraint of equal thresholds corresponds to an assumption that there was no bias toward choosing one color over another. Given evidence of substantial
overlap in estimates from the pilot study when separate thresholds were assumed, this assumption was also deemed reasonable. I therefore specified a single design matrix $X_{\kappa}$ and column vector $\beta_{\kappa}$ of coefficients for both thresholds. Using the exemplar set of observations $\{y_1, y_2, y_3, y_4\}$, then

$$\kappa_A = \kappa_B = X_{\kappa} \beta_{\kappa} = \begin{bmatrix} y_1 \\ 1 \\ y_2 \\ 1 \\ y_3 \\ 1 \\ y_4 \\ 1 \end{bmatrix} \begin{bmatrix} \kappa \end{bmatrix}.$$ 

Additional columns and coefficients will be added in later sections.

Finally, there is a non-decision component (also known as a residual latency), $\tau$. For simplicity I did not decompose the residual latency any further. To aid presentational clarity, let $\theta = \{\kappa_A, \xi_A, \kappa_B, \xi_B, \tau\}$, the set of parameters for the racing diffusion process. Then for the decision process these parameters can be represented as

$$\theta_D = \{X_{\kappa} \beta_{\kappa}, X_{\xi_A} \beta_{\xi_A}, X_{\kappa} \beta_{\kappa}, X_{\xi_B} \beta_{\xi_B}, \tau\}, \quad (3.3)$$

and a subject’s response times and choices were then distributed as

$$(t, y) \sim f_D(t, y|\theta_D). \quad (3.4)$$

### 3.1.4 The Interference Process

As summarized by Proctor et al. (2013), many theories for the Simon effect posit that a response is driven by two different processing routes: a indirect, more controlled route versus a direct, more automatic route. The automatic route results in activation of the response that corresponds spatially to the stimulus, which can facilitate performance in the congruent condition but interfere with performance in the incongruent condition. I still hypothesized that this facilitation and interference
that produce the Simon effect would be (partially) reflected in a relative ordering of the drift rates by congruency and correctness. However, as I demonstrated in section 3.1.1, a single racing diffusion process that possesses this ordering of drift rates is unable to predict the Simon effect. I therefore posited that a certain percentage of responses were determined solely by the automatic route, and I modeled this by adding another racing diffusion process, which I dub the interference process. For this process, the spatial location of the stimulus drives the drift rates. I specified two drift rates corresponding to choosing the correct spatial location: $\xi_{CS}$ and $\xi_{IS}$. The subscripts $C$ and $I$ once again refer to the congruent and incongruent conditions respectively, while the subscript $S$ indicates that these specifications refer to the interference process.

Let $X_{\xi_A S}$ and $X_{\xi_B S}$ be the design matrices for the parameters $\xi_A$ and $\xi_B$ respectively. Furthermore, let $\beta_{\xi_A S}$ and $\beta_{\xi_B S}$ be column vectors of coefficients. To provide an example of the dummy coding I used for these design matrices, I refer again to the exemplar set of observations $\{y_1, y_2, y_3, y_4\}$. Assume that for these observations a subject would always make a left key-press to indicate the choice “red”. A subject’s choice was encoded in terms of color. Therefore, the correct spatial location of observations $y_1$ and $y_4$ is represented by the “red” choice, because $y_1$ comes from a congruent trial with a red stimulus, whereas $y_4$ comes from an incongruent trial with an orange stimulus (i.e. the orange stimulus was shown on the left side of the screen). For observations $y_2$ and $y_3$ the correct spatial location is represented by the “orange” choice. Therefore, I needed a coding scheme such that for observations $y_4$, $\xi_A = \xi_I S$, 


while for $y_2$, $\xi_B = \xi_{CS}$. The dummy code presented below achieves this goal.

$$\xi_A = X_{\kappa_{AS}} \beta_{\kappa_{AS}} = \begin{bmatrix} C_{left} & C_{right} & I_{left} & I_{right} \\ y_1 & 1 & 0 & 0 & 0 \\ y_2 & 0 & 1 & 0 & 0 \\ y_3 & 0 & 0 & 0 & 1 \\ y_4 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \xi_{CS} \\ 0.01 \\ \xi_{IS} \\ 0.01 \end{bmatrix},$$

and

$$\xi_B = X_{\kappa_{BS}} \beta_{\kappa_{BS}} = \begin{bmatrix} C_{left} & C_{right} & I_{left} & I_{right} \\ y_1 & 1 & 0 & 0 & 0 \\ y_2 & 0 & 1 & 0 & 0 \\ y_3 & 0 & 0 & 0 & 1 \\ y_4 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.01 \\ \xi_{CS} \\ 0.01 \\ \xi_{IS} \end{bmatrix}.$$

As can be seen, the drift rates for the incorrect spatial locations across both conditions were fixed to an extremely low value (0.01) for convenience and model identifiability. This constraint corresponds to the assumption that subjects would virtually always be able to correctly identify the spatial location, though their speed in doing so may vary. A mixture model in which there was a single accumulation process for spatial location was explored, but brief examinations suggested poor fit to the pilot study data, so that model was not considered any further.

I constrained the threshold parameters for the interference process to equal the threshold specified for the decision process. Therefore, the decomposition of $\kappa_A$ and $\kappa_B$ for the decision process in section 3.1.3 applies for the interference process as well. This formulation assumes that there is no difference in how subjects set thresholds between the two different processes. Furthermore, I assumed that the interference process shared the same residual latency $\tau$ from the decision process. Therefore the set of parameters for the interference process is

$$\theta_S = \{X_{\kappa} \beta_{\kappa}, X_{\xi_{AS}} \beta_{\xi_{AS}}, X_{\kappa} \beta_{\kappa}, X_{\xi_{BS}} \beta_{\xi_{BS}}, \tau\}, \quad (3.5)$$
and a subject’s response times and choices under the interference process are distributed as

\[(t, y) \sim f_D(t, y|\theta_S).\]  

(3.6)

The inclusion of the interference process allows the model to predict the Simon effect. Figure 3.4 provides a visual representation of the evidence accumulation for both the decision and interference process under congruent and incongruent conditions based on when a red stimulus was shown. The boxes on the left are for the decision process, while the ones on the right for the interference process. The boxes on the top are for the congruent condition, while the ones on the bottom are for the incongruent condition. The average path of the evidence accumulation for each choice is represented by an arrow, with the “red” and “orange” response accumulators colored appropriately. Well-performing subjects were expected to have higher values of \(\xi_{CR}\) and \(\xi_{IR}\) relative to \(\xi_{CW}\) and \(\xi_{IW}\). In the incongruent condition, spatial location should interfere with the decision process. If this is the case, then it is expected that \(\xi_{IR} < \xi_{CR}\) and \(\xi_{IW} > \xi_{CW}\). This ordering of drift rates in the decision process cannot predict the extremely fast errors (when a response is based on the spatial location instead of the color of a stimulus) in the incongruent condition. Instead, such responses are generated by the interference process when \(\xi_{IS}\) is extremely high. The parameters \(\rho_{CD}, \rho_{CS}, \rho_{ID},\) and \(\rho_{IS}\) govern the frequency in which responses are generated from the two processes in the different conditions. These parameters can be thought of representing the shift in attention between processes based on congruency. I have provided example percentages to represent my prediction that in the incongruent condition subjects will fail to inhibit the interference process more often.
Figure 3.4: The decision (left-hand side) and interference (right-hand side) processes in both the congruent (top) and incongruent (bottom) condition. The parameter $\kappa$ is the threshold, $\xi$ refers to the drift rate, and $\rho$ governs the frequency in which a response is generated from a particular process. The arrows represent the average path of the evidence accumulation, and their color indicates the choice to which they correspond.
Also note that for computational ease, in the following section I use an alternative formulation to the parameters $\rho_{CD}$, $\rho_{CS}$, $\rho_{ID}$, and $\rho_{IS}$.

3.2 Slow Responses

I have so far discussed the two processes that allow my model to predict the Simon effect. I now introduce a third process to account for extremely slow responses that may not arise from either the decision or the interference processes, and instead may reflect a failure of attention on the subject’s part. While such responses are from a qualitatively different process, standard sequential sampling models treat these observations as any other response, which can introduce extreme biases in parameter estimates (e.g., Ratcliff & Tuerlinckx, 2002). The data can be examined and such extremely slow responses removed, but because I was interested in sequential effects within a block, particularly post-error behavior, I did not want to throw out data on a trial to trial basis.

An alternative to removing the trials is to explicitly model the process producing these extreme responses (Craigmile, Peruggia, & Van Zandt, 2010), and treat the data as a mixture of these separate processes. I introduced a third process to include in my mixture model for the data (without this component the model could not fit the data). I make few theoretical assumptions about the mechanisms underlying this final process. For convenience, I assumed that response times and choices arising from this process were independent from each other. I allowed the response times to follow a log-normal distribution with mean parameter $\mu$ and standard deviation fixed to 0.3, and I allowed the choices to follow a Bernoulli distribution with the probability parameter fixed to 0.5. Under this model, the joint distribution for a response time
\( t \) and choice \( y \) is
\[
(t, y) \sim f_A(t, y | \mu),
\] (3.7)
where
\[
f_A(t, y | \mu) = \frac{1}{0.3 \sqrt{2\pi}} \exp \left( -\frac{(\ln t - \mu)^2}{2(0.3)^2} \right) \times (0.5)^y (1 - 0.5)^{(1-y) I(t, y)}. \] (3.8)

Here \( I(t, y) \) is an indicator function that equals 1 for \( t > 0 \) and \( y \in \{0, 1\} \) and 0 otherwise. I chose to fix the standard deviation to 0.3 because the number of responses (ideally) coming from this process should be small, making it difficult to estimate the standard deviation parameter. The choice of 0.3 produces a balance such that the distribution has a tail long enough to encompass very slow responses up to 4 seconds, but the variability is not so wide as to interfere with the other processes.

### 3.2.1 The Mixture Model

The full mixture model has the probability density function:
\[
f(t, y | \theta) = \rho_D f_D(t, y | \theta_D) + \rho_S f_S(t, y | \theta_S) + \rho_A f_A(t, y | \mu),
\] (3.9)
where \( \theta \) refers to the set of all parameters currently defined. The mixing probabilities \( \rho_m \) were constrained so that \( \sum_{m \in \{D,S,A\}} \rho_m = 1 \) and \( \rho_m \geq 0 \). To impose these restrictions during estimation, I made use of two reparameterizations, drawn from work by Bock (1972) and Thissen and Steinberg (1984). First, a row vector of two parameters \([\lambda_D, \lambda_S]\) with support from \((-\infty, \infty)\) is multiplied by a 2 by 3 transformation matrix
\[
T = \begin{bmatrix}
\frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\
-\frac{1}{3} & \frac{2}{3} & -\frac{1}{3}
\end{bmatrix}
\]
thereby creating a row vector \( \delta \) of 3 elements constrained to sum to 1. Next, the values of \( \delta \) are constrained to lie between 0 and 1 using the multivariate logistic transform:
\[
\rho_m = \frac{\exp \delta_m}{\sum_{i=1}^{3} \exp \delta_i}.
\]
I also decomposed parameters \( \lambda_D \) and \( \lambda_A \) into the weighted sum of a set of covariates and coefficients. I let \( \lambda_D = X_{\lambda_D} \beta_{\lambda_D} \) and \( \lambda_S = X_{\lambda_S} \beta_{\lambda_S} \). To describe the dummy coding used for these design matrices, I again refer to the exemplar set of responses \( \{y_1, y_2, y_3, y_4\} \). Then

\[
X_{\lambda_D} \beta_{\lambda_D} = \begin{bmatrix} C & I \\ y_1 & 1 & 0 \\ y_2 & 1 & 0 \\ y_3 & 0 & 1 \\ y_4 & 0 & 1 \end{bmatrix} \begin{bmatrix} \lambda_{CD} \\ \lambda_{ID} \end{bmatrix},
\]

and

\[
X_{\lambda_S} \beta_{\lambda_S} = \begin{bmatrix} C & C_S & I & I_S \\ y_1 & 1 & -1 & 0 & 0 \\ y_2 & 1 & -1 & 0 & 0 \\ y_3 & 0 & 0 & 1 & -1 \\ y_4 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \lambda_{CD} \\ \lambda_{CS} \\ \lambda_{ID} \\ \lambda_{IS} \end{bmatrix}.
\]

With this code scheme, when the support of \( \lambda_{CS} \) and \( \lambda_{IS} \) is restricted to include positive values only, the coefficients decrement the values of \( \lambda_{CD} \) and \( \lambda_{ID} \). This scheme ensures that the probabilities for the interference effect will always be less than those for the decision. This constraint is especially useful for the congruent condition, because without it the correspondence between response and stimulus location leave little information for the model to untangle the decision and interference processes based on the covariates alone. Next, in the following sections I describe the additional covariates and parameters I incorporated into my model to examine my research goals.

### 3.3 The Simon Task and Motivation

I now describe how I accounted for differences in model parameters based on the lack of payoffs in the neutral condition versus the payoffs heavily penalizing errors in the motivation condition.
Unequal payoffs. I expected that model parameters would be impacted by payoffs penalizing errors more strongly in the motivation condition in two ways (assuming the manipulation was effective). Suppose that for the set of exemplar observations \{y_1, y_2, y_3, y_4\}, the final pair of observations comes from the motivation condition. Then in this condition, a subject may become more cautious, increasing their accuracy at the cost of speed. This corresponds to an increase in the response thresholds. To examine this, I added an additional column to the design matrix $X_\kappa$:

$$X_\kappa \beta_\kappa = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \begin{bmatrix} C_A & C_B & I_A & I_B & M \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \kappa \\ \kappa M \end{bmatrix}.$$ 

Here, the coefficient $\kappa_M$ is an increment to the threshold for the payoff condition.

Furthermore, subjects may also focus more and be less distracted during the motivation condition, leading to an increase in accuracy and speed due to more efficient sampling of information from the stimulus. In the model, I expect the change in speed and accuracy to be because of an increase in the drift rate for the accumulator corresponding to the correct choice, represented by the coefficient $\xi_M$. To examine this, I added an additional column to the design matrices $X_{\xi AD}$ and $X_{\xi BD}$:

$$X_{\xi AD} \beta_{\xi AD} = \begin{bmatrix} C_A & C_B & I_A & I_B & M \\ y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \begin{bmatrix} \xi_{CR} \\ \xi_{CW} \\ \xi_{IR} \\ \xi_{IW} \end{bmatrix},$$

and

$$\xi_B = X_{\xi BD} \beta_{\xi BD} = \begin{bmatrix} C_A & C_B & I_A & I_B & M \\ y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \begin{bmatrix} \xi_{CW} \\ \xi_{CR} \\ \xi_{IW} \\ \xi_{IR} \end{bmatrix}.$$
If people were more highly motivated during the payoff condition, then I predict that \( \kappa_M > 0 \) and \( \xi_M > 0 \).

There is the possibility that the motivation condition may have an impact on the drift rates for the interference process as well, though it is harder to predict in what direction. For greater flexibility in the model, I included an increment \( \xi_M S \) to drift rates corresponding to the correct spatial location based on whether a trial came from the motivation condition. I therefore also added an additional column to the design matrices \( X_{\xi AS} \) and \( X_{\xi BS} \):

\[
X_{\xi AS} \beta_{\xi AS} = \begin{bmatrix}
C_{\text{left}} & C_{\text{right}} & I_{\text{left}} & I_{\text{right}} & M \\
y_1 & 1 & 0 & 0 & 0 \\
y_2 & 0 & 1 & 0 & 0 \\
y_3 & 0 & 0 & 0 & 1 \\
y_4 & 0 & 0 & 1 & 0 \\
\end{bmatrix} \begin{bmatrix}
\xi_{CS} \\
0.01 \\
\xi_{IS} \\
0.01 \\
\xi_M \\
\end{bmatrix},
\]

and

\[
\xi_B = X_{\xi BS} \beta_{\xi BS} = \begin{bmatrix}
C_{\text{left}} & C_{\text{right}} & I_{\text{left}} & I_{\text{right}} & M \\
y_1 & 1 & 0 & 0 & 0 \\
y_2 & 0 & 1 & 0 & 0 \\
y_3 & 0 & 0 & 0 & 1 \\
y_4 & 0 & 0 & 1 & 0 \\
\end{bmatrix} \begin{bmatrix}
0.01 \\
\xi_{CS} \\
0.01 \\
\xi_{IS} \\
\xi_M \\
\end{bmatrix}.
\]

### 3.3.1 The Change in Performance over Blocks

Another important modeling consideration is how a person’s performance changes over time. As noted by Dutilh, van Ravenzwaaij, et al. (2012) changes in a person’s motivation and concentration over the course of the experiment can confound measures of post-error slowing. Dutilh, van Ravenzwaaij, et al. provide a hypothetical example of a subject who starts off fast, accurate, and engages in no post-error slowing. However, he or she then becomes fatigued over the course of the experiment, producing slower, more error-prone responses by the end of the session (still without
any post-error slowing). When separately pooling the responses following an error and a correct response, most of the correct responses would come from early in the session, when the subject was faster, while the error responses would come from later in the session, when a subject was slower. Without taking into account the subject’s fatigue, it would falsely appear that the subject had slowed down following an error, but sped up following a correct response. Conversely, aggregating data for a subject who is cautious and slow early on, but gets more careless later in the study, producing faster, more error-prone responses, can mask the actual presence of post-error slowing. Therefore, if I do not take into account the effects of fatigue, motivation, or practice effects, I could potentially undermine my examination of post-error behavior.

Modeling Changes over Blocks

A straightforward way to model changes in performance over time is to account for the change in parameters across blocks. This is a simplification, as it has been suggested, based on large undulations seen in time series plots of response time data, that a subject’s motivation and concentration wax and wane from trial to trial (e.g. Craigmile et al., 2010), and such information could be lost by pooling over blocks. I therefore allowed the parameters for threshold and the drift rates for correct choices to change based on blocks and the experimental session.

While there are many ways to model trends over blocks, such as autoregressive structures (Craigmile et al., 2010), for simplicity I let parameters exhibit a linear increase or decrease from a starting point to an end point. Again, while curvilinear relations are more likely, assuming linear trends will still allow a flexible approach that can (at least partially) account for fatigue or practice effects. Additionally, subjects can show substantially variability from session to session (such as having a better
understanding of instructions by the second session, leading to an improvement in performance). Therefore, I allowed the degree of change from block to block to vary from session to session, and I included an overall increment to drift rates based on the session.

To better describe the coding scheme I used in the design matrices to implement this design, I expand my exemplar set of observations to include an additional set of four observations, \( \{y_5, y_6, y_7, y_8\} \). For presentational purposes, I assumed the following designations:

\[
\begin{bmatrix}
\begin{array}{cc}
\text{Block} & \text{Session} \\
y_1 & 1 & 1 \\
y_2 & 1 & 1 \\
y_3 & 2 & 1 \\
y_4 & 2 & 1 \\
y_5 & 15 & 2 \\
y_6 & 15 & 2 \\
y_7 & 16 & 2 \\
y_8 & 16 & 2 \\
\end{array}
\end{bmatrix}
\]

**Changes in threshold over blocks.** To implement this design, I added two additional columns to the design matrix \( X_\kappa \). Let \( n_{bl} \) be the total number of experimental blocks a subject completes, and \( b_i \) refer to the block number for trial \( i \). Then for the \( i^{th} \) trial, values in the first new column equalled \( (b_i - 1)/(n_{bl} - 1) \) for the first session and 0 otherwise. Values in the second new column equalled \( (b_i - 1)/(n_{bl} - 1) \) for the second session and 0 otherwise. Using the exemplar observations, the design matrix
and column vector of coefficients are then

$$X_{\kappa \beta_{\kappa}} = \begin{bmatrix} M & b_1 & b_2 \\ y_1 & 1 & 0 & 0 & 0 \\ y_2 & 1 & 0 & 0 & 0 \\ y_3 & 1 & 1 & 0.07 & 0 \\ y_4 & 1 & 1 & 0.07 & 0 \\ y_5 & 1 & 0 & 0 & 0.93 \\ y_6 & 1 & 0 & 0 & 0.93 \\ y_7 & 1 & 1 & 0.1 & 1 \\ y_8 & 1 & 1 & 0.1 & 1 \end{bmatrix} \begin{bmatrix} \kappa \\ \kappa_M \\ \kappa_{b_1} \\ \kappa_{b_2} \end{bmatrix}.$$ 

The column headings $b_1$ and $b_2$ indicate the covariates by block for sessions 1 and 2 respectively. The coefficient $\kappa$ now represents the threshold for the first block, which changes according to a weighted increment $\kappa_{b_1}$ or $\kappa_{b_2}$ (depending on the session), the weight increasing from 0 to 1 based on the current block (meaning the increment has the largest impact on the final block).

**Changes in drift rate over blocks.** I also assumed that the drift rates for correct choices in both the main decision process and interference process would change as a function of the current block and session. For the main decision process, I added three additional columns to the matrices $X_{\xi_{AD}}$ and $X_{\xi_{BD}}$. Using the exemplar set of observations, the design matrices and column vectors are therefore

$$X_{\xi_{AD}\beta_{\xi_{AD}}} = \begin{bmatrix} C_A & C_B & I_A & I_B & M & 2 & b_1 & b_2 \\ y_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ y_2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ y_3 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ y_4 & 0 & 0 & 0 & 1 & 0 & 0 & 0.07 \\ y_5 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0.93 \\ y_6 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ y_7 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ y_8 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \xi_{CR} \\ \xi_{CW} \\ \xi_{IR} \\ \xi_{IW} \\ \xi_M \\ \xi_2 \\ \xi_{b_1} \\ \xi_{b_2} \end{bmatrix},$$
and

\[
X_{\xi_{BD}}\beta_{\xi_{BD}} = \begin{bmatrix}
C_A & C_B & I_A & I_B & M & 2 & b_1 & b_2 \\
y_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
y_2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
y_3 & 0 & 0 & 1 & 0 & 0 & 0 & 0.07 \\
y_4 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
y_5 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
y_6 & 0 & 1 & 0 & 0 & 0 & 1 & 0.93 \\
y_7 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
y_8 & 0 & 0 & 0 & 1 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\xi_{CW} \\
\xi_{CR} \\
\xi_{IW} \\
\xi_{IR} \\
\xi_M \\
\xi_S \\
\xi_{b_1} \\
\xi_{b_2}
\end{bmatrix}.
\]

The pair of coefficients \(\xi_{CR}\) and \(\xi_{IR}\) now represent the drift rates for correct responses at the start of the first session. In turn, \(\xi_2\) represents the change in the initial level of the drift rates for correct choices for the second session. Finally, the coefficients \(\xi_{b_1}\) and \(\xi_{b_2}\) capture the gradual increase or decrease in the correct drift rates over blocks.

For the interference process, while I did not consider gradual changes based on blocks in drift rate, I did allow an increment based on the session to which a trial belonged. I added another column (labeled \(S_2\)) to the design matrices \(X_{\xi_{AS}}\) and \(X_{\xi_{BS}}\):

\[
X_{\xi_{AS}}\beta_{\xi_{AS}} = \begin{bmatrix}
C_{\text{left}} & C_{\text{right}} & I_{\text{left}} & I_{\text{right}} & M & S_2 \\
y_1 & 1 & 0 & 0 & 0 & 0 \\
y_2 & 0 & 1 & 0 & 0 & 0 \\
y_3 & 0 & 0 & 0 & 1 & 0 \\
y_4 & 0 & 0 & 1 & 0 & 1 \\
y_5 & 1 & 0 & 0 & 0 & 1 \\
y_6 & 0 & 1 & 0 & 0 & 0 \\
y_7 & 0 & 0 & 0 & 0 & 1 \\
y_8 & 0 & 0 & 1 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\xi_{CS} \\
\xi_{IS} \\
\xi_M \\
\xi_{S_2}
\end{bmatrix},
\]

and

\[
\xi_B = X_{\xi_{BS}}\beta_{\xi_{BS}} = \begin{bmatrix}
C_{\text{left}} & C_{\text{right}} & I_{\text{left}} & I_{\text{right}} & M & S_2 \\
y_1 & 1 & 0 & 0 & 0 & 0 \\
y_2 & 0 & 1 & 0 & 0 & 0 \\
y_3 & 0 & 0 & 0 & 1 & 1 \\
y_4 & 0 & 0 & 1 & 0 & 0 \\
y_5 & 1 & 0 & 0 & 0 & 0 \\
y_6 & 0 & 1 & 0 & 0 & 1 \\
y_7 & 0 & 0 & 0 & 1 & 1 \\
y_8 & 0 & 0 & 1 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
0.01 \\
\xi_{CS} \\
0.01 \\
\xi_{IS} \\
\xi_M \\
\xi_{S_2}
\end{bmatrix}.
\]
The parameter $\xi_{s_2}$ represents the increment to the drift rate for the spatial location based on the second session.

### 3.3.2 Post-error Adjustments

In the previous sections, I discussed how I attempted to account for the Simon effect, motivation, and trends over blocks in the data. By accounting for these effects, I will be able to better capture the variability in the data attributable to post-error adjustments. My model allows for multiple types of errors. First, subjects can make fast errors due to a failure to inhibit the results of automatic processing of the spatial location of a stimulus. Additionally, subjects can successfully inhibit this automatic processing, yet still fail to correctly process the color of the stimulus. Finally, subjects may make slow guesses due to lapses of attention over the course of the experiment, resulting in a response that involved little to no processing of the stimulus.

The previous studies by Pieters et al. (2007), Schrijvers et al. (2010), and (Tops et al., 2013) examined post-error slowing with the assumption that it was a behavioral marker of greater cognitive control following errors. As noted by Dutilh, Vandekerckhove, et al. (2012), in the sequential sampling framework this assumption of increased cognitive control maps onto an increase in response thresholds for trials following an error. In this case, more evidence must accumulate, resulting in slower, more accurate responses. While there are other mechanisms by which a subject may slow down following an error, such as a reorientation effect due to surprise (Notebaert et al., 2009), I focused on an increase in response caution in order to directly test the assumptions made by Pieters et al. (2007), Schrijvers et al. (2010), and (Tops et al., 2013). However, following slower errors (such as an attentional failure), it seems
unreasonable to expect subjects to slow down. Instead, subjects may return to their original rapid decision process. The combination of slowing following faster errors but no slowing following slower errors could potentially attenuate measurements of the degree of post-error slowing, which in turn could reduce the power in detecting associations with perfectionist characteristics. To check this possibility, I tested whether people made two types of post-error adjustments. Finally I took into account the work of Luu et al. (2000), who found that an aggregate measure of post-error slowing decreased over blocks. This finding is consistent with the observation of an attenuated ERN amplitude (the ERP component associated with more cautious responding) in later blocks compared to earlier blocks (Luu et al., 2000; Pailing & Segalowitz, 2004; Tops et al., 2013).

**Hypotheses Regarding Errors**

I have three hypotheses regarding errors:

- following fast error responses, subjects will be more likely slow down and be more accurate;

- following slow error responses, subjects will not engage in post-error slowing; and

- subjects will show more post-error slowing in early blocks compared to later blocks.

I now describe the covariate structures that will allow me to examine these hypotheses. I specified four different models: first, a model that assumes no shift in thresholds for post-error slowing; second, a model that assumes a shift in thresholds following...
any error; third, a model that assumes shifts in thresholds following fast errors only; finally, a model that assumes shifts in thresholds following fast errors that can differ based on the first or second half of a session.

**The first model.** To test whether post-error adjustments are present in the data, it is first important to have a reference model that assumes no post-error slowing. The first model therefore was fit with the covariate structure and coefficients as defined above according to Equation 3.9, with no additions for post-error adjustment.

**The second model.** If subjects become more cautious following an error, their response thresholds should shift upward. I can test this by adding an increment to the response thresholds based on whether a trial follows an error or not. Therefore, for the second model, I added a column to $X_\kappa$ whose values equalled 1 if in the previous trial the subject committed an error. Furthermore, the coefficient $\kappa_{P_1}$, representing the shift in the threshold following an error, was added to $\beta_\kappa$.

**The third model.** There are several ways to specify a model to examine whether subjects only engage in post-error slowing following fast errors. A straightforward method is to, in contrast to the second model, add a new column to $X_\kappa$ whose values equalled 1 on trials that followed an error, but only for errors which were faster than the response preceding the mistake. Let $E$ denote the trial in which an error was made. Let $t_E$ be the response time associated with the error. Then if the response time for the trial preceding the error is greater than the response time for the error ($t_{E-1} > t_E$), then the value in the new column for $X_\kappa$ for trial $E + 1$ equalled 1. The coefficient included in $\beta_\kappa$ is instead $\kappa_{P_2}$, representing a shift in response caution that
only occurs following fast errors, regardless of the degree of difference in response
time between the error and preceding response.

**The fourth model.** Examining whether post-error adjustments change over time
can be a challenging problem when the frequency of errors are low. Unfortunately,
the pilot study indicated that this is the case with my experiment, as subjects’ error
rates only ranged from about 5% to 20%. In fact, in some cases subjects completed
entire blocks with no errors. Therefore, I instead examined differences in post-error
adjustments over the first and second half of a session, to ensure a sufficient number
of errors to estimate the parameters.

Again, I specify an alternative addition to $X_\kappa$ compared to the previous model
variants. In this case, I add two additional columns, whose values once again equal 1
on trials that followed an error, specifically errors which were faster than the response
preceding the mistake. However, for the first column, values could only equal 1 for
trials in the first half of a session, while for the second column values could only equal
1 for the second half of the session. Two new rows were included in the coefficient
matrix $\beta_\kappa$, with the parameters $\kappa_{P,1}$ and $\kappa_{P,2}$.

### 3.3.3 Summary of model parameters

The models I have defined have a large number of parameters. For ease of un-
derstanding, I present all the freely estimated model parameters in Table 3.1, along
with a brief interpretation of each and the model variants that make use of a given
parameter.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Interpretation</th>
<th>Model variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>Intercept</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\kappa_M$</td>
<td>Shift based on payoffs</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\kappa_{b1}$</td>
<td>Gradual shift for session 1</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\kappa_{b2}$</td>
<td>Gradual shift for session 2</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\kappa_P$</td>
<td>Post-error shift</td>
<td>2</td>
</tr>
<tr>
<td>$\kappa_{P1}$</td>
<td>Shift following fast errors</td>
<td>3</td>
</tr>
<tr>
<td>$\kappa_{Pb1}$</td>
<td>Shift following fast errors for first half</td>
<td>4</td>
</tr>
<tr>
<td>$\kappa_{Pb2}$</td>
<td>Shift following fast errors for second half</td>
<td>4</td>
</tr>
<tr>
<td>$\xi_{CR}$</td>
<td>Drift rate for correct congruent choices</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\xi_{CW}$</td>
<td>Drift rate for incorrect congruent choices</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\xi_{IR}$</td>
<td>Drift rate for correct incongruent choices</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\xi_{IW}$</td>
<td>Drift rate for incorrect incongruent choices</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\xi_{M}$</td>
<td>Shift based on payoffs</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\xi_{2}$</td>
<td>Shift for the second session</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\xi_{B1}$</td>
<td>Gradual shift for session 1</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\xi_{B2}$</td>
<td>Gradual shift for session 2</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\xi_{CS}$</td>
<td>Drift rate for congruent spatial location</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\xi_{IS}$</td>
<td>Drift rate for incongruent spatial location</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\xi_{MS}$</td>
<td>Shift based on payoffs</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\xi_{S2}$</td>
<td>Shift for the second session</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\tau$</td>
<td>Non-decision component</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Location parameter for slow responses</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\lambda_{CD}$</td>
<td>Probability of congruent main decision process</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\lambda_{ID}$</td>
<td>Probability of incongruent main decision process</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\lambda_{CS}$</td>
<td>Probability of congruent interference process</td>
<td>1 - 4</td>
</tr>
<tr>
<td>$\lambda_{IS}$</td>
<td>Probability of incongruent interference process</td>
<td>1 - 4</td>
</tr>
</tbody>
</table>

Table 3.1: List of free parameters for all model variants.
3.3.4 Impact of Perfectionism

I now describe how I expect perfectionist characteristics to impact post-error slowing. Previous studies found little relation between measures of perfectionism and post-error slowing. Schrijvers et al. (2010) found that, though there was significant post-error slowing among subjects, there was no relationship between post-error slowing and any of their administered inventories. Furthermore, in the study by Tops et al. (2013) an aggregate measure of post-error slowing had a non-significant negative correlation with the “Concerns over mistakes” subscale of the FMPS. Instead, post-error slowing correlated positively with a self report measure of arousal and negatively with the late Pe, which Tops et al. (2013) argue represents conscious awareness of errors. One of my hypotheses is that, by using the diffusion race model to better parse out underlying sources of variability, we will be able to detect potential relationships between perfectionism measures and post-error slowing.

To examine these potential associations between perfectionistic characteristics and the underlying mechanisms specified in the above models, I examined how well sum scores on the “Doubts about actions” and the “Concerns over mistakes” subscales predicted point estimates for coefficients taken from the best-fitting model variant. The set of parameters that I was most interested in was the threshold parameter $\kappa$, and the shifts in the threshold following an error ($\kappa_{P_1}, \kappa_{P_2}, \text{or } \kappa_{P_b1}$ and $\kappa_{P_b2}$ depending on the model).

I first hypothesized that higher scores on the higher values on the “Doubts about actions” and the “Concerns over mistakes” subscales will predict a higher value for the intercept term $\kappa$. This is driven by the findings of Pieters et al. (2007), in which anorexia patients (a group well-known for increased perfectionism) displayed a more
controlled, cautious response style during a flanker task. In turn, the correlation between the “Doubts about actions” subscale and the ERN amplitude that Schrijvers et al. (2010) found suggests that increased response caution should correspond with greater uncertainty over actions (another intuitive hypothesis). In this case, higher scores on the “Doubts about actions” subscale should predict higher thresholds, and a higher threshold following an error.

The lack of correlation between scores from the “Concerns over mistakes” subscale of the FMPS and the ERN (Schrijvers et al., 2010; Tops et al., 2013) suggests that this subscale will have likely have no relation with increased response caution following an error. One interesting possibility is that the “Concerns over mistakes” subscale will account for the change in post-error slowing over time. For instance, Tops et al. hypothesized that subjects who scored higher on the “Concerns over mistakes” subscale engaged in emotional regulation to avoid the negative affect associated with mistakes, leading to a dampened ERN. If this is the case, I predict a reduction in response caution as the experiment progresses. This suggests that higher scores on the “Concerns over mistakes” subscale result in a larger difference between the shift in threshold following an error for the first and second half of a session. Note that if the fourth model does not fit well, that indicates that there is insufficient information from the behavioral data to test this hypothesis. I included the sum scores from the negative affect items of the state subscale of the STAI as a control, especially in light of the finding of Luu et al. (2000) in which subjects with higher negative affect had higher initial post-error slowing, but showed a faster decline compared to the low negative affect group. Hence, anxiety could play a more important role in post-error behavior and in any observed declines in this behavior.
It is also possible that perfectionist characteristics will predict a change in performance (or a lack of change) based on motivation. I therefore examined the association between the coefficients based on the motivation condition \( \kappa_M \) and \( \xi_M \) and the “Doubts about actions” and “Concerns over mistakes” subscales. Pailing and Segalowitz (2004) found that the increase in the ERN based on incentivized conditions was strongly positively correlated with higher scores on the personality factor of Neuroticism. Both the “Doubts about actions” and “Concerns over mistakes” subscales show moderate correlations with the Neuroticism factor (Frost et al., 1990). Therefore I predicted that those who scored higher both subscales should show a higher positive shift in the coefficients \( \kappa_M \) and \( \xi_M \). As before, I included the negative affect items from the STAI subscale as a control.

As an exploration of possible associations, I also included the coefficients meant to account for global changes in performance \( \kappa_{B_1}, \kappa_{B_2}, \xi_2, \xi_{B_1}, \xi_{B_2} \). It is likely that the set of coefficients under examination will be highly correlated with each other, making it also beneficial to include these coefficients in the analysis to avoid ignoring such correlations. I examined the association between these coefficients and the inventory scores, but I did not make any specific hypotheses regarding the direction of the associations. This variables served as important controls, but as the degree of fatigue and learning could vary to a large degree from subject to subject, making it difficult to predict any specific pattern of results.

I also included an intercept, which would represent the base value of each coefficient when inventory scores were equal to 0. In the actual analysis, I standardized the inventory scores, but left the coefficient variables in their original scale. Note that here the value 0 here would represent the average score on an inventory subscale. I
had certain predictions regarding the intercepts for the coefficients $\kappa_M$, $\zeta_M$, and either $\kappa_{P_1}$, $\kappa_{P_2}$, or $\kappa_{P_b}$ and $\kappa_{P_{b2}}$ depending on the model. For these coefficients, I predicted that their associated intercept term would be greater than 0, indicating a significant increase in response caution during the motivation condition and following errors, as well as a significant increase in attention during the motivation condition, leading to higher drift rates.

To test for these predictions, I fit a multivariate normal model to the point estimates of interest (because there were multiple point estimates for each subject, this can be thought of as a repeated measures design). Let $\theta^j$ be a column vector containing the point estimates of interest for the $j^{th}$ subject. Then

$$
\theta^j \sim \text{MVN}(\mu^j_\theta, \Sigma),
$$

(3.10)

where $\mu^j_\theta$ is a column vector of means for subject $j$, and $\Sigma$ is the associated covariance matrix. The mean vector can be decomposed as

$$
\mu_\theta = X^j \beta.
$$

(3.11)

Here $\beta = [\beta_0, \beta_{CoM}, \beta_{DaA}, \beta_{SN}]^T$, a column vector with the regression coefficients for the intercept, “Doubts about actions”, “Concerns over mistakes”, and STAI subscales (the latter three represented by the subscripts $CoM$, $DaA$, and $SN$). In turn, $X^j = [1, x^j_{CoM}, x^j_{DaA}, x^j_{SN}]$, the $j^{th}$ row of a design matrix containing in the final three columns the standardized sum scores for the subscales for all subjects. No constraints were imposed on the covariance matrix $\Sigma$, with all parameters freely estimated.
Chapter 4: Analyses

I first review the basics of Bayesian estimation, discuss the particular algorithms and estimation methods I relied on for fitting my model, and note some general advantages of the Bayesian approach. I then detail the methods I used to check model fit, compare models, and assess differences and associations between my parameters and inventory measures.

4.1 Bayesian Estimation

In a frequentist framework, parameters are assumed to be fixed within a group of experimental trials, and inference is based on the sample space created if the experiment were to be replicated many times. By contrast, the Bayesian framework treats parameters as random variables along with the data, and inferences are based on the probability distributions for the parameters after some data has been observed (Turner & Van Zandt, 2012). The probability distributions of the parameters given observed data are known as posterior distributions. The posterior quantifies the uncertainty around the model’s parameters subsequent to observing the data and given one’s prior knowledge about a parameter before the data were collected. The posterior distribution can be calculated using Bayes’ rule:

\[ p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}, \]  

(4.1)
where $D$ refers to the observed data and $\theta$ to the set of model parameters; $p(D|\theta)$ is the likelihood, the probability of observing the current data given the model parameters; $\pi(\theta)$ is the prior distribution of the parameters; and $p(D)$ is the marginal probability of the data, acting as a normalizing constant because it does not explicitly depend on the model of interest. For discrete data

$$p(D) = \sum_{\theta_i \in \Theta} p(D|\theta_i)\pi(\theta_i)$$

and for continuous data

$$p(D) = \int_{\theta \subset \Theta} p(D|\theta)\pi(\theta)d\theta.$$  

The normalizing constant is often intractable, which in the early days of Bayesian analysis imposed severe constraints on applications. However, advances such as a computer-driven MCMC sampling methodology (Gamerman & Lopes, 2006) can circumvent this issue. This methodology involves the construction of a Markov chain that has the posterior as its equilibrium distribution. The unknown parameters are drawn from proposal distributions and adjusted in iterative steps to approximate the posterior; because each new value drawn depends only on the previously drawn value, they form a Markov chain. For my project, I used the Metropolis-Hastings (MH) algorithm (Metropolis & Ulam, 1949; Hastings, 1970) to implement the Markov chains. In the following section, I briefly review how the MH algorithm works.

4.2 MH Algorithm

As reviewed by Gelman et al. (2013), starting with some initial set of values for the parameters, on each iteration $t$ the MH algorithm proceeds by generating a new set of candidate values $\theta^*$ from a proposal distribution conditioned on the current
parameters $\theta_t$. The candidate values are accepted with probability $\alpha$, where

$$\alpha = \min \left( 1, \frac{g(\theta^*) \pi(\theta^*) q(\theta_t | \theta^*)}{g(\theta_t) \pi(\theta_t) q(\theta^* | \theta_t)} \right).$$

(4.2)

In the equation, $g(\theta)$ is the likelihood (how likely the data are given the model parameters), $\pi(\theta)$ is the prior for the model parameters, and $q(a|b)$ is the density or kernel function for the proposal distribution used to generate the candidate values. In this case, the normalizing constant cancels out, thereby solving the issue of the intractable denominator. Moreover, if a symmetric proposal distribution is used, the kernel functions also cancel, further simplifying the equation. Typically, the proposal distribution is Gaussian with a mean equal to $\theta_t$ and a variance functioning as a tuning parameter, the value of which fixed by the experimenter to give the best rate of convergence towards the equilibrium distribution. The final step in the algorithm involves generating a random deviate from a uniform distribution with range 0 to 1. This number is compared to $\alpha$, and if $\alpha$ exceeds this value, the candidate values are accepted and we set $\theta_{t+1} = \theta^*$; if not, we set $\theta_{t+1} = \theta_t$. This process is repeated for $m$ iterations, generating a chain $\theta = (\theta_1, \theta_2, ..., \theta_m)$ that can be taken as a sample of the posterior.

The MH algorithm is a powerful tool but it can be an inefficient method of sampling for many types of models. The efficiency of the sampler depends in large part on the tuning parameters used in the proposal distributions when generating new candidate values. For instance, consider a Gaussian proposal distribution. In this case, the tuning parameter is the variance of the distribution. Setting the variance too high results in most candidate values being too extreme and therefore rejected. If the variance is too low, most candidate values will be accepted, but the range of values will be too small to properly traverse the parameter space. In both cases, the
resulting samples will not properly represent the target distribution. Therefore, when using the MH algorithm, the tuning parameters must be selected to strike a balance between these two extremes. When parameters are correlated, such a balance can be difficult to determine. This holds true especially for the sequential sampling model I used to address questions about perfectionism. Sequential sampling models can be difficult to fit with conventional MCMC methods. For instance, when fitting the diffusion model using Bayesian methods, Vandekerckhove, Tuerlinckx, and Lee (2011) comment that real life analyses could take hours if not days, due to the computational burden and inefficiency of the samplers.

4.3 Differential Evolution MCMC

To address the issue of inefficient samplers, I used differential evolution MCMC (DE-MCMC; Ter Braak, 2006; Turner, Sederberg, Brown, & Steyvers, 2013), which generates more efficient proposals for candidate values. The DE-MCMC algorithm uses multiple interacting chains, and the difference between the current states for a selection of chains is used to generate proposals for other chains. Assuming K chains (where K must be equal or greater than four), a new proposal for the $k^{th}$ chain is generated by

$$\theta^* = \theta_k + \gamma(\theta_m - \theta_n) + \epsilon,$$  (4.3)

for $m \neq n \neq k$. Here, $\theta_k$ is the current state of the $k^{th}$ chain, while $\theta_m$ and $\theta_n$ are the current states for two other chains selected at random from the entire set of chains. The term $\epsilon$ is a small amount of random noise added to avoid degeneracy problems. Given that $U(a, b)$ refers to the uniform distribution with lower and upper bounds $a$ and $b$, then typically $\epsilon \sim U(-b, +b)$ with $b = 0.001$ or some similarly small
value. The final term $\gamma$ is a positive tuning parameter that controls the magnitude of the change.

Unlike the original MH algorithm, the DE-MCMC method does not require a separate tuning parameter for each parameter to be specified, making it simpler. Typical settings for the tuning parameter are $\gamma = 2.38\sqrt{2d}$ where $d$ is the number of dimensions of the parameter space (Ter Braak, 2006). After the new proposal has been generated, it is accepted or rejected using the MH probability defined in equation 4.2. Turner et al. (2013) demonstrated the efficacy of the DE-MCMC method by using it to find posterior estimates for a hierarchical LBA model fit to data from a moving dots task. Turner et al. were unable to apply the standard MH algorithm to these data because of the highly correlated parameter space. Therefore, the DE-MCMC method is an important tool for efficient parameter estimation when fitting choice and response time data from human subjects.

4.4 Migration Step

Mixture distributions have always posed a particular challenge for estimation routines, because the distributions are often multimodal in nature, making it easy for the routine to get stuck in particular regions of the parameter space. Furthermore, as noted by Turner et al. (2013), individual chains in the DE-MCMC can perform quite badly when they are poorly initialized as well. Because I generated the starting points of the chains randomly, the risk of bad starting values was higher than desired. I therefore implemented a migration step. The use of the migration step to deal with outlier chains was originally proposed by Hu and Tsui (2005), and recently
has seen expanded use in modeling efforts within Psychology (Turner & Sederberg, 2012; Turner et al., 2013).

As described by Turner et al. (2013), the basic idea in a migration step is to propose a jump from the current state of one chain to another. This proposal can be expanded to include multiple chains, such that several chains can be swapped in cyclical fashion. The algorithm presented by Turner et al. involves first sampling a number \( \eta \) uniformly from the set \( k = \{1, 2, \ldots, K\} \), where \( k \) is the total number of chains. Next, \( \eta \) numbers are sampled without replacement from \( k \), forming a group set \( G = \{G_1, G_2, \ldots, G_\eta\} \). The current state of each chain \( \theta_{G_i} \) indexed by \( G \) can then be swapped in the aforementioned cyclic fashion:

\[
\{\theta_{G_1}, \theta_{G_2}, \ldots, \theta_{G_{\eta-1}}, \theta_{G_\eta}\} \rightarrow \{\theta_{G_\eta}, \theta_{G_1}, \ldots, \theta_{G_{\eta-2}}, \theta_{G_{\eta-1}}\}.
\] (4.4)

Turner et al. note that doing this swap deterministically does not solve the issue of poorly performing outlier chains. They therefore introduce a small amount of noise to each proposal:

\[
\theta_{G_\eta}^* = \theta_{G_\eta} + \epsilon_\eta, \\
\theta_{G_1}^* = \theta_{G_1} + \epsilon_1, \\
\vdots \\
\theta_{G_{\eta-1}}^* = \theta_{G_{\eta-1}} + \epsilon_{\eta-1}.
\]

Whether a proposal is swapped or not is then determined probabilistically, using the MH algorithm. The authors note that, if values of \( \epsilon_i \) are sampled from a symmetric distribution, the proposal probabilities do not have to be included in the MH step. I therefore sampled these values from a uniform distribution with a range of -0.1 to 0.1.
4.5 Advantages of the Bayesian Method

There are other advantages to working in a Bayesian framework. First, Bayesian methods allow the incorporation of prior information. In other words, we can specify our uncertainty regarding model parameters before any data are collected. Priors can be useful because they allow information from past research or pilot-studies to contribute to the current analyses, or provide a means to incorporate information from experts. Priors help reduce the variance and uncertainty around parameter estimates (Kruschke, 2011). Moreover, from a model-fitting perspective, priors can be a useful tool to constrain model parameters estimates to lie within reasonable bounds, thereby improving the stability of the estimation process.

For example, consider the parameter for response caution (i.e. boundary separation) in sequential sampling models. Technically, this parameter can take on any positive value up to infinity. However, for extremely large values of boundary separation, especially given a small rate of evidence accumulation, the likelihood of the process terminating at a boundary becomes so small that for all intents and purposes it equals zero. Because a response time of infinity is empirically impossible, a prior on the response caution parameter can be set to ensure that such extreme values are impossible during the model estimation process. However, it is important to avoid priors that overpower the influence of the observed data. This is less of an issue as sample size grows, because with more and more data the likelihood comes to dominate, and the influence of the prior becomes negligible (Gelman, Carlin, Stern, & Rubin, 2009).

Another attractive feature of the Bayesian framework is that it allows sequential updating (Lee & Wagenmakers, 2014). In other words, if additional data from a new
set of subjects are collected following a previous study, the posterior of the older study can be used as the prior for the analysis of the new data, allowing us to update our uncertainty on the entire dataset. An important consequence of sequential updating is that stopping rules for experiments have less impact on statistical assumptions. In a frequentist framework, the underlying statistical assumptions can be drastically different based on the sampling scheme used (Wagenmakers, 2007). For instance, optional stopping and examining the data before the total sample size has been achieved can inflate Type I error (Francis, 2012). Bayesian methods do not have such issues.

Another benefit of the Bayesian framework is the intuitive interpretation of the results. For example, as noted by Gelman et al. (2009), a probability or confidence interval for an unknown quantity of interest in the Bayesian framework does in fact represent the desired probability of containing the unknown quantity. In contrast, for the frequentist framework, a confidence interval must be interpreted in relation to a sequence of similar inferences that in theory might be made repeatedly (which, as anyone who has taught an introductory statistics course would know, is an unintuitive concept for many people to grasp).

4.6 Priors

Working within a Bayesian framework, I must specify priors for each of the estimated coefficients (see Table 3.1). To my knowledge, no previous study has conducted a Bayesian analysis using the diffusion race model, necessitating some care about the choice of priors. My choices were motivated by a combination of wanting to impose weak constraints on some coefficients while wanting to use informed priors on others.
to ensure greater stability in estimation. When imposing weak constraints, I used uniform distributions with restricted ranges. I used truncated normal distributions for the more informed priors. Conservative assumptions were made for the priors placed on the shift coefficients (e.g., $\kappa_M, \kappa_{P_1}, \xi_M$, etc.), with the mean of the truncated normal set to 0 and a standard deviation of 0.5. The lower and upper boundaries for the truncated normal priors reflected the constraints I placed on the proposals for a given coefficient.

A very informed prior was placed on the coefficient $\mu$, as the (ideally) low occurrence of slow responses meant there was little information in the data to help estimate this parameter. Furthermore, the informed prior on this coefficient helped stabilize estimates by ensuring that the log-normal process only accounted for a small proportion of the variability in the data. In the following, $N(m, s)$ refers to the normal distribution with mean $m$ and standard deviation $s$, and $TN(m, s, a, b)$ refers to the normal distribution with mean $m$ and standard deviation $s$, truncated at lower and upper bounds $a$ and $b$. The same priors were used for each subject. The priors for my threshold coefficients were

$$\{\kappa_M, \kappa_{B_1}, \kappa_{B_2}, \kappa_{P_1}, \kappa_{P_2}, \kappa_{P_3}, \kappa_{P_4}\} \sim TN(0, 0.5, -2, 2),$$

and $\kappa \sim U(0, 20)$.

The priors for the drift rates for the main decision process were

$$\{\xi_{CR}, \xi_{IR}\} \sim U(0, 20),$$

$$\{\xi_{CW}, \xi_{IW}\} \sim TN(2, 5, 0, 5),$$

and $\{\xi_M, \xi_2, \xi_{B_1}, \xi_{B_2}\} \sim TN(0, 5, -5, 5)$. 

105
The priors for the drift rates for the interference process were

\[
\{\xi_{CS}, \xi_{IS}\} \sim U(0, 20),
\]

and \(\{\xi_{MS}, \xi_{S2}\} \sim TN(0, .5, -2, 2)\).

Finally, the priors for the ancillary processes and transformed probabilities were

\[
\tau \sim U(0, \min(t)),
\]

\[
\mu \sim N(0, 0.1),
\]

\[
\{\lambda_{CD}, \lambda_{ID}\} \sim N(4, 1),
\]

and \(\{\lambda_{CS}, \lambda_{IS}\} \sim TN(2, .5, 0, 10)\).

### 4.7 Estimation routine

I now describe the general structure of my estimation routine, which I implemented using the statistical software R (R Core Team, 2014), and in particular the R to C++ interface package Rcpp (Eddelbuettel & François, 2011). First, I specified the total number of chains to use in the DE-MCMC algorithm to generate new proposals, as well as the total number of iterations to run for each chain. I set the number of chains equal to the number of parameters (both fixed and free) multiplied by 15. Hence, 360 chains were created for the first model, 375 chains for the second and third models, and 390 chains for the fourth. I ran 1500 iterations in which I drew samples for each chain (except in the case of the 15th subject, for which I ran 2500 iterations, because of initial difficulty with convergence). The tuning parameter \(\gamma\) for the DE-MCMC algorithm was set to 0.3, while \(\epsilon\) was fixed to 0.001, as these values produced better acceptance rates for the more complex models.
Starting values. I created disperse starting values for the coefficients of each chain by drawing random values from the uniform distribution. Table 4.1 gives the lower and upper limit used for each coefficient when drawing values from the uniform distribution. The choice of starting values reflected in part my hypotheses regarding coefficients (for instance, the range for the drift rate of the interference process in the incongruent condition was set to a higher interval relative to the other drift rates). In other cases, the starting values were chosen to ensure that the initial likelihood values were defined. Hence, several of the coefficients leading to shifts in parameters (e.g. \( \kappa_M \)) were given a small range between -0.01 and 0.01 to avoid producing negative values. Similarly, the range for the residual latency \( \tau \) was restricted to lie below 90% of the lowest response time provided by a subject.

Sampling. I now describe the sampling process I used for each iteration and chain. At the start of an iteration, a migration step occurred with a 5% probability (determined by generating a uniform random number between 0 and 1 and determining if it was less than 0.05). I then looped through each chain. For greater computational speed, I ran this loop in parallel using the OpenMP library (OpenMP Architecture Review Board, 2014). For each chain, I generated new proposals for the coefficients using the DE-MCMC algorithm. To further improve acceptance rates, for several coefficients proposal values were restricted to only fall within a certain range. Table 4.1 also provides the boundaries between which proposal values had to fall. Additionally, I made use of block sampling: I split the coefficients into two sets based on pilot runs. Then for each set, I determined whether to accept or reject the new proposals with a MH step. This method helped improve acceptance rates by reducing the risk that the
<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Range</th>
<th>Boundary</th>
<th>Block</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>[0.5, 2.0]</td>
<td>-2.0, 2.0</td>
<td>1</td>
</tr>
<tr>
<td>$\kappa_M$</td>
<td>[-0.1, 0.1]</td>
<td>-2.0, 2.0</td>
<td>1</td>
</tr>
<tr>
<td>$\kappa_{B_1}$</td>
<td>[-0.1, 0.1]</td>
<td>-2.0, 2.0</td>
<td>1</td>
</tr>
<tr>
<td>$\kappa_{B_2}$</td>
<td>[-0.1, 0.1]</td>
<td>-2.0, 2.0</td>
<td>1</td>
</tr>
<tr>
<td>$\kappa_{P_1}$</td>
<td>[0.0, 0.25]</td>
<td>-2.0, 2.0</td>
<td>1</td>
</tr>
<tr>
<td>$\kappa_{P_2}$</td>
<td>[0.0, 0.25]</td>
<td>-2.0, 2.0</td>
<td>1</td>
</tr>
<tr>
<td>$\kappa_{P_{11}}$</td>
<td>[0.0, 0.25]</td>
<td>-2.0, 2.0</td>
<td>1</td>
</tr>
<tr>
<td>$\kappa_{P_{12}}$</td>
<td>[0.0, 0.25]</td>
<td>-2.0, 2.0</td>
<td>1</td>
</tr>
<tr>
<td>$\xi_{CR}$</td>
<td>[3.0, 6.0]</td>
<td>0.0, $\infty$</td>
<td>2</td>
</tr>
<tr>
<td>$\xi_{CW}$</td>
<td>[0.5, 2.0]</td>
<td>0.0, 5.0</td>
<td>1</td>
</tr>
<tr>
<td>$\xi_{IR}$</td>
<td>[3.0, 6.0]</td>
<td>0.0, $\infty$</td>
<td>2</td>
</tr>
<tr>
<td>$\xi_{IW}$</td>
<td>[0.5, 2.0]</td>
<td>0.0, 5.0</td>
<td>1</td>
</tr>
<tr>
<td>$\xi_{M}$</td>
<td>[-0.1, 0.1]</td>
<td>-5.0, 5.0</td>
<td>2</td>
</tr>
<tr>
<td>$\xi_2$</td>
<td>[-0.1, 0.1]</td>
<td>-5.0, 5.0</td>
<td>2</td>
</tr>
<tr>
<td>$\xi_{B_1}$</td>
<td>[-0.1, 0.1]</td>
<td>-5.0, 5.0</td>
<td>2</td>
</tr>
<tr>
<td>$\xi_{B_2}$</td>
<td>[-0.1, 0.1]</td>
<td>-5.0, 5.0</td>
<td>2</td>
</tr>
<tr>
<td>$\xi_{CS}$</td>
<td>[0.5, 2.0]</td>
<td>0.0, $\infty$</td>
<td>2</td>
</tr>
<tr>
<td>$\xi_{IS}$</td>
<td>[5.0, 7.0]</td>
<td>0.0, $\infty$</td>
<td>1</td>
</tr>
<tr>
<td>$\xi_{MS}$</td>
<td>[-0.1, 0.1]</td>
<td>-5.0, 5.0</td>
<td>1</td>
</tr>
<tr>
<td>$\xi_{S_2}$</td>
<td>[-0.1, 0.1]</td>
<td>-5.0, 5.0</td>
<td>2</td>
</tr>
<tr>
<td>$\tau$</td>
<td>[0.001, (0.9)min(t)]</td>
<td>0.001, min(t)</td>
<td>1</td>
</tr>
<tr>
<td>$\mu$</td>
<td>[-0.25, 0.25]</td>
<td>$-\infty$, $\infty$</td>
<td>2</td>
</tr>
<tr>
<td>$\lambda_{CD}$</td>
<td>[2.0, 4.0]</td>
<td>$-\infty$, $\infty$</td>
<td>2</td>
</tr>
<tr>
<td>$\lambda_{ID}$</td>
<td>[2.0, 4.0]</td>
<td>$-\infty$, $\infty$</td>
<td>1</td>
</tr>
<tr>
<td>$\lambda_{CS}$</td>
<td>[0.0, 2.0]</td>
<td>0.0, $\infty$</td>
<td>2</td>
</tr>
<tr>
<td>$\lambda_{IS}$</td>
<td>[0.0, 2.0]</td>
<td>0.0, $\infty$</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4.1: A list of the lower and upper limits of the range used to generate random starting values from a uniform distribution, the boundaries between which each proposal fell, and the sampling block for each coefficient.
entire set of proposals would be rejected because of one poorly performing proposal. Table 4.1 reports how different coefficients were assigned to the two sampling blocks. When the algorithm finished looping through all iterations, the samples for each chain and whether a proposal was rejected or accepted for a given iteration were stored. The data from each subject was fit separately.

4.8 Convergence

When using MCMC methods to approximate the posterior, after the samples have been drawn, the next important step is to evaluate the convergence of the chains. For early iterations, the Markov chain usually has not reached its stationary distribution and therefore the values it generates will be poor estimates of the posterior distribution (Kruschke, 2011). This issue can be addressed by specifying a burn-in period, in which values from the chain are discarded so that they do not overly influence the final estimate of the posterior. For each model variant and subject, the chains were visually inspected and a burn-in of the initial 500, 700, or 900 samples was taken. The set of samples for each chain produced by the DE-MCMC algorithm were then combined, thereby providing a large number of observations with which to approximate the posterior.

However, while the DE-MCMC algorithm allows rapid convergence even when parameters are highly correlated with each other, samples in the chain can still be highly autocorrelated, which can result in some values being over-represented while others are under-represented in the posterior estimate (Kruschke, 2011). To address this, I thinned my combined chains, using only every 20\textsuperscript{th} step of the chain to ensure sufficient lag between points such that autocorrelations were no longer an issue. To
double check that the thinned chains (following burn-in) were indeed stationary, I calculated the Geweke diagnostic (Geweke, 1992) for each chain. The Geweke diagnostic compares the mean of two non-overlapping parts of the chain (I used the first 10% and the last 50%), and assesses whether the means differ significantly using z-scores.

4.9 Model Comparisons and Fit

After I obtained estimates of the posteriors and ensured that the chains for these estimates had converged, I then determined the best fitting model of the four variants, and evaluated how well the model predicted the observed data.

4.9.1 Model Comparisons

I relied on the Watanabe-Akaike information criterion (WAIC; Watanabe, 2010) to compare model performance. Vehtari and Gelman (2014) provide a useful tutorial on how to calculate the WAIC. Vehtari and Gelman note that the WAIC, like other information criteria, is intended as a measure of predictive accuracy of a model. Specifically, the WAIC is meant as an approximation to a model’s expected log pointwise predictive density for a new dataset. This density is based on the posterior predictive density for a new observation $\tilde{y}_i$, which is

$$p(\tilde{y}_i|y) = \int_{\theta \in \Theta} p(\tilde{y}_i|\theta)p(\theta|y)d\theta.$$  \hspace{1cm} (4.5)

Here, $\theta$ is a set of parameters, and $y$ the original set of observed data, with $y = \{y_1, ..., y_N\}$. In turn, $p(\theta|y)$ is the posterior distribution, determined given a set of priors $\pi(\theta)$. Finally, $p(\tilde{y}_i|\theta)$ is the likelihood of the new observation given $\theta$. The authors let $f_\theta(y)$ refer to the distribution of the true data-generating process that depends on $\theta$. The expected log pointwise predictive density ($elpd$) for a new dataset
\( \bar{y} = \{ \bar{y}_1, ..., \bar{y}_N \} \) is then
\[
\sum_{i=1}^{N_n} \mathbb{E}_{f_i}(\log p(\bar{y}_i|y)).
\] (4.6)

Because \( \theta \) is unknown in practice, the WAIC or cross-validation methods must be used to approximate the elpd.

In the calculation of the elpd in practice, Vehtari and Gelman define a useful quantity, the log pointwise predictive density (lppd) which is based on the observed data. This quantity can be approximated using draws from a MCMC sample of the posterior distribution, with an individual draw represented as \( \theta^s \) (where \( s = 1, ..., S \) given \( S \) total draws). The computation is
\[
lppd = \sum_{i=1}^{N} \left( \frac{1}{S} \sum_{s=1}^{S} p(y_i|\theta^s) \right). \]
(4.7)

However, the authors emphasize that the lppd is an overestimate of the elpd for future data. Therefore, a correction term is needed:
\[
elpd = \hat{lppd} - \hat{\rho}.
\] (4.8)

Here, \( \hat{\rho} \) is an estimate of the effective number of parameters, which can be calculated using the posterior variance of the log predictive density for each data point \( y_i \). Let \( V(a) \) refer to the calculation of the sample variance, \( \frac{1}{S-1} \sum_{s=1}^{S} (a_s - \bar{a})^2 \). Then
\[
\hat{\rho} = \sum_{i=1}^{N} V(\log p(y_i|\theta)).
\] (4.9)

Vehtari and Gelman then note that elpd can be transformed to a deviance scale (in which smaller values indicate less deviance and better predictive accuracy) using the formulation
\[
WAIC = -2\hat{elpd}.
\] (4.10)
This is the information criterion I calculated for each model for comparison purposes. I calculated the WAIC for each subject and model separately. To get an overall estimate of the WAIC each model, I also pooled the values of $y_i$ and $\theta^*$ for each subject and calculated the WAIC on this pooled sample. Vehtari and Gelman also provide a formula to obtain an estimate of the standard error for the WAIC. They define a new term

$$e\hat{lpd}_i = \log \left( \frac{1}{S} \sum_{s=1}^{S} p(y_i|\theta^*) \right) - (V(p(y_i|\theta))) \right),$$

where $e\hat{lpd} = \sum_{i=1}^{N} e\hat{lpd}_i$. Then for the set $e\hat{lpd}_{all} = \{e\hat{lpd}_1, ..., e\hat{lpd}_N\}$ the standard error is the square root of $NV(e\hat{lpd}_{all})$.

4.9.2 Posterior Predictive Checks

To further examine how well the models fit the data, I carried out posterior predictive checks. Posterior predictive checks involve simulating data from the model using parameter values randomly drawn each time from the estimated posteriors. This simulated data may then be compared with the observed data by plotting the data on the posterior predictions (Gelman & Shalizi, 2013; Gelman et al., 2009). If the predicted data approximate the observed data with no systematic discrepancies, this suggests good fit. A numerical posterior predictive check can also be performed (Gelman et al., 2009) by first specifying test quantities (such as credible intervals) for the observed data.

To create the posterior predictive checks for my models I simulated 1000 sets of data with equivalent sample sizes and stimulus presentations for each subject. I then calculated the relevant test quantities (see below) for each separate simulation, and determined 95% credible intervals by finding the 0.025 and 0.975 quantiles for the set.
of test quantities. I relied on several different test quantities to assess the performance of several aspects of my model. To examine how well my model fit subjects’ response times conditioned on congruency and accuracy, I used a quantile-probability plot (Ratcliff, 2001) to compare the observed data and the predicted quantities from the model. To determine how well my model accounted for trends, I calculated the predicted quantities for the accuracy and overall response time quantiles for each block.

4.10 Evaluating Point Estimates

Following the determination of the best fitting model, I calculated point estimates by estimating the mode of each posterior distribution. I did this by taking the highest value returned by the `density()` function in R. After I obtained point estimates of the coefficients for each subject, I fit the multivariate normal model described earlier using the R interface to Stan (Stan Development Team, 2014a), a general purpose package for Bayesian analysis. Stan uses another variant of MCMC sampling known as Hamiltonian Monte Carlo (HMC; Neal, 2011). HMC, similar to DE-MCMC, allows much faster convergence to a stationary distribution through much more efficient proposals. This is achieved by using the gradient of the log probability function. Unlike DE-MCMC, HMC is difficult to tune, but Stan solves this problem using the No-U-turn sampler (NUTS; Hoffman & Gelman, 2011), which can automatically adapt the number of steps during sampling such that no tuning is necessary for the user. Once I obtained samples from the posterior, I once again evaluated convergence using the Geweke diagnostic criterion (Geweke, 1992).
In fitting the model, I had to specify priors for the regression coefficients describing the association between the coefficients for the cognitive model and the inventory scores, as well as the covariance matrix. In the former case, I specified uninformed priors for the coefficients relating to the intercepts, using a normal distribution with a mean of 0 and a standard deviation of 5. For the coefficients corresponding to the associations with subscales, I was concerned that the large number of coefficients I was examining would increase the risk of a false-positive finding. Therefore, I used more informed normal priors that were tightly clustered around 0 with a standard deviation of 0.5. Such priors necessitate a greater degree of information indicating the presence of an association. This more conservative approach helped mitigate concerns regarding multiple comparisons.

In specifying a relatively uninformed prior for the covariance matrix $\Sigma$, I followed the recommendations given in the manual for Stan (Stan Development Team, 2014b).

Let $\Omega$ be a correlation matrix, and $D$ refer to a diagonal matrix based on the vector $\eta$ of variable scales. The covariance matrix can then be decomposed as

$$\Sigma = D\Omega D.$$ 

This decomposition also aids in numerical stability. In specifying priors, I used a weakly informative prior on $\eta$, a Cauchy distribution with a location parameter set to 0 and a scale parameter set to 5. For $\Omega$, I used a $LJK$ distribution (Lewandowski, Kurowicka, & Joe, 2009) with a shape parameter equal to 2. As noted in the Stan manual, for the $LJK$ distribution, as the scale parameter increases, the prior increasingly concentrates around the unit correlation matrix, such that the prior favors less correlation among the components. The Stan manual provides example code for such implementations.
Another important consideration when fitting the multivariate normal model to the point estimates was the fact that not all subjects completed both sessions. Seven subjects had only completed session 1, and for one subject (as will be discussed later), data were available only for the second session. In the former case, point estimates were not available for the coefficients $\kappa_{B_2}$ and $\xi_{B_2}$. In the latter case, point estimates were not available for $\kappa_{B_1}$ and $\xi_{B_1}$. In both cases, because there was only a single session, point estimates were also not available for $\xi_2$ and $\xi_{S_2}$. Therefore, the appropriate parameters were excluded from the mean vector and covariance matrix when fitting these particular subjects.
Chapter 5: Results

In this chapter, I present the findings of my experiments, including my model fit, and the associations between coefficients and inventory scores. I first discuss my pre-analysis data processing.

5.1 Preprocessing

Before I analyzed the data, it was necessary to remove some sessions and blocks of data based on several criteria. This removal of data was necessary for three reasons. First, subjects would occasionally fail to make a response within 4 seconds, resulting in a timeout response: no choice was registered and the experiment proceeded on to the next trial. These trials were removed. Second, sometimes subjects would respond extremely rapidly, far faster than possible for any normal decision process (for instance, responses made within 5 ms of the stimulus presentation). Such responses are problematic because they prevent accurate estimation of residual latencies and can induce biases in parameter estimates, and so they also were removed. Finally, subjects’ performance could occasionally drop to chance levels, often with some combination of extremely rapid and timeout responses. Such data are problematic because they contain no information to inform parameter estimates, in addition to the potential to bias estimates of trend over time, and so they were removed.
Timeout responses. Because I was interested in sequential effects, post-error behavior in particular, I decided to not remove responses on a trial to trial basis. Trial to trial exclusion of specific responses could potentially mask the presence of sequential effects. Instead, when there was a low occurrence of timeout responses, the data could be recoded with a random choice in which the selection of a color was equally likely in the few cases in which such responses did happen. In this way, the slow response component of the model would account for these timeout responses, and analysis could proceed normally. However, blocks (and sessions) with a high occurrence of such timeout responses would result in a large amount of missing data (and possibly the worst kind of missing data, missing not at random), potentially biasing parameter estimates. Therefore, if 10% or more of the trials in single session for a subject were timeout responses, that session was removed. Furthermore, if 10% or more of the trials in single block of data for a subject were timeout responses, that block was removed.

Rapid responses. No mechanism was present in the model for extremely rapid responses. Even a single extremely rapid observation could prove deeply problematic. A response was deemed extremely rapid if it fell below 50 ms, and in such cases, the entire block was removed.

Chance performance. A useful criterion to evaluate a subject’s performance is the $d'$ statistic (Green & Swets, 1966), taken from Signal Detection Theory (SDT). Using one the simplest models from SDT, we can assume the existence of a single dimension in a subject’s psychological space that can be used to discriminate between stimulus types (whether a stimulus was red or orange in this case). Two distributions,
Table 5.1: Sessions and blocks removed based on percentage of timeouts, fast responses, and extremely poor performance.
one representing the perception of the color red, and the other the perception of the
color orange, lie on this psychological dimension. The simplest model assumes that
these distributions are normally distributed with a common variance 1, and that the
mean of the distribution for the color orange equals 0. The mean of the distribution
for perceiving red is then estimated as

\[ d' = \Phi^{-1}(H) - \Phi^{-1}(FA), \]  

where \( \Phi^{-1}(\cdot) \) refers to the inverse of the standard normal cumulative distribution,
\( H \) refers to the proportion of “Hits” (correct identification of a red stimulus), and
\( FA \) refers to the proportion of “False alarms” (when an orange stimulus is identified
as red). Higher values of \( d' \) indicate better performance, while a value of 0 indicates
chance performance. The upper boundary of the 90% confidence interval around such
chance performance is 0.3. This value was used as the cutoff criterion for a block, and
I discarded a subject’s block if the \( d' \) statistic computed for it fell below this cutoff. I
report all sessions and blocks that were removed for each subject as well as the reason
in Table 5.1. Blocks numbered higher than 16 come from the second session.

5.2 Exploratory Analysis

As a basis for comparison against my later discussion of the benefits of a modeling
approach, I now present several figures detailing the aggregate performance of the
subjects on several different test statistics. In the creation of the quantile-probability
plots used in this section, the 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles, as well
as mean accuracy per congruency condition, were calculated for each subject. I then
averaged across subjects to produce the figures. The ellipses represent two standard
errors above and below the mean for the test statistics, calculated by estimating the
standard deviation across subjects and then dividing by the square root of the sample size (i.e. 29).

**The Simon effect.** As noted in section 3.1, the Simon effect is a typical pattern of data found in the Simon task in which subjects make faster errors and slower correct responses in the incongruent condition. By contrast, for the congruent condition, subjects instead have slower errors but faster correct responses. Figure 5.1 presents a quantile-probability plot showing correct and error responses based on whether the condition was congruent or incongruent. The red ellipses denote the location of the median for the response times. The figure shows that the average median response time for correct responses in the incongruent condition is over two standard errors higher compared to the median response time for incorrect response times in the same condition. Furthermore, the median response time for correct responses in the congruent condition is over two standard errors lower compared to correct responses in the incongruent condition, but this pattern is reversed for error responses. Finally, responses times exhibit the characteristic right skew seen in human performance data (e.g. Luce, 1986). My study provides a successful replication of the typical effects seen in the Simon task.

**The effect of payoffs.** Figure 5.2 presents a quantile-probability plot for correct and error response times based on whether subjects earned money in a given block of trials. The 0.1, 0.3, 0.5, 0.7, and 0.9 quantiles (averaged across subjects) are shown again. The circles denote the quantiles from conditions in which the subjects earned money, whereas the crosses denote the conditions in which subjects earned no money. Ellipses once again represent two standard errors above and below the mean quantile.
Figure 5.1: Quantile-probability plot showing circles and crosses (representing the congruent and incongruent conditions, respectively) for the observed 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles averaged over subjects on the y-axis and probability of a correct or incorrect response on the x-axis. Ellipses represent plus or minus two standard errors.
Figure 5.2: Quantile-probability plot showing circles and crosses (representing the payoffs and no payoffs conditions, respectively) for the observed 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles averaged over subjects on the y-axis and probability of a correct or incorrect response on the x-axis. Ellipses represent plus or minus two standard errors.
Quantiles for the two conditions are extremely close, with substantial overlap of the standard errors, indicating no noticeable effect of the motivation manipulation on either accuracy or response times.

**Trends over blocks of trials.** Figure 5.3 details the changes over block for the average accuracy across subjects on the left, and the changes over block for average response time quantiles across subjects on the right. The shaded areas denote two standard errors above and below the mean accuracy or mean quantile for each individual block (i.e. an approximate 95% confidence interval). The dashed line shows the average accuracy across all blocks on the left, and the average median response time across all blocks on the right. While there seems to a slight downward trend in accuracy over blocks (suggesting a fatigue effect), the 95% confidence intervals for each individual block always overlap with the average accuracy across all blocks. A researcher relying solely on this aggregate measure of trend in accuracy would therefore conclude that there is no fatigue or practice effects present in the data. The response time quantiles show slightly slower responses with wider variability in early blocks, and slightly faster responses with less variability for the final blocks, suggesting a practice effect. The 95% confidence intervals for the average median response time per individual block does not overlap with the overall median across all blocks for these initial and final blocks. This suggests a significant, albeit small, practice effect over time.

**Post-error slowing** One of the main goals of my project was to evaluate post-error behavior and relate that to perfectionism. I now report two different measures of post-error slowing, the traditional measure computed by comparing the mean response
Figure 5.3: The left plot shows changes in average accuracy by block, while the right plot shows changes in the averaged 0.1, 0.3, 0.5, 0.7, and 0.9 quantiles by block. Shaded regions denote two standard errors above and below the mean, providing approximate 95% confidence intervals. The dashed lines represent the average accuracy (on the left) and average median response time (on the right) across all blocks.
Figure 5.4: Averages for two different measures of post-error slowing. The traditional measure is the mean of correct response times following an error minus the mean of correct response times following a correct choice. The robust measure is the mean of the difference in response times for trials following an error and trials preceding an error. Lines represent two standard errors above and below the mean.
times following an error and following a correct response, as well as the more robust
measure proposed by Dutilh, van Ravenzwaaij, et al. (2012) which calculates the
mean of the difference in response times for the trials following an error and the trials
preceding an error. For the traditional measure, only correct post-error responses are
typically considered (Dutilh, van Ravenzwaaij, et al., 2012), and as such I followed
suit when calculating the measure here (though this ignores the potentially interesting
information that repeat errors can provide).

In calculating the robust measure of post-error slowing, I excluded errors that
occurred on the first trial, because there are no preceding trials in which to include
in the calculation. Figure 5.4 reports the average for both the robust and traditional
measure across subjects, with the error bars representing two standard errors above
and below the mean. Both measures indicate that, on average, subjects slowed by
about 70 ms following an error. Because the standard errors for both measures do
not overlap with zero, this indicates the degree of slowing is significant. Note that
both measures are very similar, which makes sense given the previous indication that
there was little change in accuracy and response times over blocks and sessions on
average.

These aggregate measures of performance suggest that subjects reliably produced
the Simon effect, were unaffected by the motivation manipulation, and showed little
fatigue over the course of a session. Furthermore, these statistics suggest that subjects
exhibited a mild practice effect, but most importantly, when controlling for global
trends in the data, exhibited no post-error slowing. However, as I noted in my
introduction, aggregate measures of performance can be misleading in these tasks. I
now present the results of fitting my four model variants to the data. I first discuss
the convergence of each model, report model comparisons using the WAIC for both individual subjects and pooled data, and then I present posterior predictive checks for select subjects.

5.3 Convergence

Examination of the chains following estimation indicated that while most chains converged successfully, for almost every subject there was a small proportion of poorly performing chains. Therefore, I inspected each set of chains per coefficient, subject, and model, and then established which of the three burn in periods (500, 700, or 900 samples) to use, and which chains to exclude. Figure 5.5 provides an example (taken from the estimates of $\xi_{CS}$ for Subject 12 and the second model) of the method I used to identify poorly performing chains to remove. The solid vertical line represents the burn-in I chose in this case. In most cases, it was easy to identify which chains had converged, and these were used to establish a upper and lower limit (the red lines in the figure). Chains whose mean value over an interval of at least 100 iterations fell outside these limits were eliminated.

As can be seen in the figure, about three or four chains had values that ended up at much higher values relative to the other chains. These chains were excluded. Following removal of bad chains, I pooled the samples from each set of chains for a coefficient together, and then thinned to every 20th observation (to reduce autocorrelations). Even after thinning, posterior samples typically consisted of between 12,000 to 17,000 observations, providing a great deal of reassurance that these were stable estimates of the posterior distributions. Table 5.2 reports the minimum and maximum values of the Geweke diagnostic criterion (Geweke, 1992) across all coefficients.
for the final set of posterior samples of each subject and model variant. Because this criterion is a z-score, extreme values indicate a larger difference in the mean of the initial 10% of the posterior sample relative to the mean for the final 50%. As can be seen in the table, all values are very close to 0, indicating good convergence.

The occurrence of poorly performing chains allows some insight into the stability of the parameter estimations. Across all subjects and relative to other coefficients, $\xi_{CS}$ and $\lambda_{CD}$ typically had the most occurrences of poorly performing chains, suggesting that these coefficients were the least stable and most difficult to estimate for the model. Furthermore, two subjects had estimates across all four model variants that were unstable compared to the other subjects. Subject 3 had a larger number of chains that got trapped, producing bimodal posterior estimates. For the sake of stable point estimates, I eliminated chains whose means over intervals of 100 iterations extended beyond the mode for the smaller peak of the bimodal distribution. Posterior predictive checks indicated that such removals did not affect model predictions negatively.

Subject 15 had estimates that took much longer to converge, which as noted before necessitated extending the number of iterations to 2500 instead of 1500. Subject 15 also had many more poorly performing chains. However, these chains were easily identified for specific coefficients (e.g., values of $\mu$ centered at -1.0, indicating that the log-normal process is centered around response times that are too fast to represent attentional failures). I therefore was still able to identify which chains had converged for Subject 15. Furthermore, even after thinning, the posterior sample for Subject 15 still consisted of over 4,000 observations. Later inspection of this subject’s pattern of responses revealed a pattern that was extremely difficult for the model to fit, thereby accounting for the instabilities in estimation.
Figure 5.5: Example showing the visual inspection process used to establish the burn in period and identify poorly performing chains. The solid vertical line denotes the chosen burn-in period. The dashed red lines denote the interval used to identify poorly performing chains. Chains that lie outside this interval for over 100 iterations were removed.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.10, 0.10</td>
<td>-0.08, 0.10</td>
<td>-0.09, 0.05</td>
<td>-0.07, 0.09</td>
</tr>
<tr>
<td>2</td>
<td>-0.22, 0.23</td>
<td>-0.10, 0.11</td>
<td>-0.15, 0.16</td>
<td>-0.13, 0.13</td>
</tr>
<tr>
<td>3</td>
<td>-0.14, 0.11</td>
<td>-0.10, 0.08</td>
<td>-0.07, 0.08</td>
<td>-0.13, 0.13</td>
</tr>
<tr>
<td>4</td>
<td>-0.13, 0.14</td>
<td>-0.11, 0.09</td>
<td>-0.08, 0.09</td>
<td>-0.08, 0.07</td>
</tr>
<tr>
<td>5</td>
<td>-0.05, 0.10</td>
<td>-0.06, 0.10</td>
<td>-0.10, 0.09</td>
<td>-0.08, 0.11</td>
</tr>
<tr>
<td>6</td>
<td>-0.06, 0.10</td>
<td>-0.06, 0.05</td>
<td>-0.08, 0.09</td>
<td>-0.06, 0.04</td>
</tr>
<tr>
<td>7</td>
<td>-0.02, 0.09</td>
<td>-0.08, 0.08</td>
<td>-0.04, 0.05</td>
<td>-0.07, 0.07</td>
</tr>
<tr>
<td>8</td>
<td>-0.04, 0.03</td>
<td>-0.06, 0.04</td>
<td>-0.08, 0.08</td>
<td>-0.05, 0.04</td>
</tr>
<tr>
<td>9</td>
<td>-0.04, 0.03</td>
<td>-0.13, 0.09</td>
<td>-0.04, 0.07</td>
<td>-0.10, 0.12</td>
</tr>
<tr>
<td>10</td>
<td>-0.11, 0.11</td>
<td>-0.04, 0.04</td>
<td>-0.05, 0.06</td>
<td>-0.05, 0.09</td>
</tr>
<tr>
<td>11</td>
<td>-0.22, 0.21</td>
<td>-0.14, 0.15</td>
<td>-0.12, 0.15</td>
<td>-0.12, 0.12</td>
</tr>
<tr>
<td>12</td>
<td>-0.08, 0.06</td>
<td>-0.07, 0.05</td>
<td>-0.07, 0.06</td>
<td>-0.07, 0.06</td>
</tr>
<tr>
<td>13</td>
<td>-0.14, 0.16</td>
<td>-0.08, 0.08</td>
<td>-0.10, 0.16</td>
<td>-0.11, 0.11</td>
</tr>
<tr>
<td>14</td>
<td>-0.05, 0.05</td>
<td>-0.04, 0.05</td>
<td>-0.06, 0.08</td>
<td>-0.06, 0.11</td>
</tr>
<tr>
<td>15</td>
<td>-0.04, 0.15</td>
<td>-0.06, 0.11</td>
<td>-0.09, 0.10</td>
<td>-0.15, 0.25</td>
</tr>
<tr>
<td>16</td>
<td>-0.09, 0.12</td>
<td>-0.12, 0.10</td>
<td>-0.15, 0.11</td>
<td>-0.12, 0.06</td>
</tr>
<tr>
<td>17</td>
<td>-0.08, 0.11</td>
<td>-0.07, 0.08</td>
<td>-0.07, 0.05</td>
<td>-0.05, 0.08</td>
</tr>
<tr>
<td>18</td>
<td>-0.09, 0.10</td>
<td>-0.11, 0.09</td>
<td>-0.05, 0.04</td>
<td>-0.05, 0.08</td>
</tr>
<tr>
<td>19</td>
<td>-0.20, 0.05</td>
<td>-0.14, 0.13</td>
<td>-0.10, 0.06</td>
<td>-0.11, 0.04</td>
</tr>
<tr>
<td>20</td>
<td>-0.08, 0.06</td>
<td>-0.05, 0.05</td>
<td>-0.06, 0.04</td>
<td>-0.08, 0.06</td>
</tr>
<tr>
<td>21</td>
<td>-0.07, 0.05</td>
<td>-0.09, 0.06</td>
<td>-0.06, 0.08</td>
<td>-0.06, 0.08</td>
</tr>
<tr>
<td>22</td>
<td>-0.07, 0.04</td>
<td>-0.06, 0.08</td>
<td>-0.05, 0.05</td>
<td>-0.08, 0.09</td>
</tr>
<tr>
<td>23</td>
<td>-0.05, 0.05</td>
<td>-0.06, 0.06</td>
<td>-0.10, 0.08</td>
<td>-0.08, 0.06</td>
</tr>
<tr>
<td>24</td>
<td>-0.04, 0.04</td>
<td>-0.06, 0.07</td>
<td>-0.08, 0.05</td>
<td>-0.05, 0.08</td>
</tr>
<tr>
<td>25</td>
<td>-0.05, 0.05</td>
<td>-0.06, 0.05</td>
<td>-0.08, 0.03</td>
<td>-0.07, 0.09</td>
</tr>
<tr>
<td>26</td>
<td>-0.06, 0.03</td>
<td>-0.12, 0.10</td>
<td>-0.04, 0.08</td>
<td>-0.07, 0.08</td>
</tr>
<tr>
<td>27</td>
<td>-0.05, 0.08</td>
<td>-0.08, 0.09</td>
<td>-0.07, 0.10</td>
<td>-0.06, 0.08</td>
</tr>
<tr>
<td>28</td>
<td>-0.05, 0.09</td>
<td>-0.07, 0.06</td>
<td>-0.10, 0.09</td>
<td>-0.08, 0.08</td>
</tr>
<tr>
<td>29</td>
<td>-0.06, 0.06</td>
<td>-0.07, 0.08</td>
<td>-0.06, 0.05</td>
<td>-0.07, 0.06</td>
</tr>
</tbody>
</table>

Table 5.2: The minimum and maximum Geweke diagnostic criterion values (z-scores) for each subject and model variant. More extreme values indicate large differences between the first 10% and the last 50% of the chain.
5.4 Model Comparisons

After obtaining large posterior samples, I then compared the four model variants by calculating the WAIC for individuals, as well as calculating a global WAIC by pooling together all individuals for each model variant. For computational ease, WAIC values were calculated using a sample of 2,000 coefficient values drawn from the posterior estimates. Table 5.3 presents the individual values of the WAIC (along with the standard error estimate in parentheses). Because the WAIC is a deviance criterion, smaller values indicate less deviance and hence better predictive accuracy of the model. The smallest WAIC for each individual and model has been bolded.

In general, Model 2 exhibits the best performance, though differences between models never exceed even one standard error. This trend is reflected in Figure 5.6, which presents the global WAIC values for each model variant, along with error bars indicating two standard errors above and below each estimate. All models with a mechanism for post-error adjustment have a lower WAIC compared to the reference model, and Model 2 has the lowest WAIC overall. Although the estimates all fall within each other’s error bars, because of the trend that Model 2 showed, as well as its lower complexity, I chose to focus on Model 2 predominantly for the rest of the analyses.

One should be very cautious about the conclusions drawn about model performance based on global measures of fit such as the WAIC. First, because of the large number of components in the model, the model can perform quite well with most components and still have one aspect that increases the deviance. For instance, well-performing subjects will have very few extremely slow responses. Even though the performance of a subject is well described by the model, the log-normal process will
<table>
<thead>
<tr>
<th>Subject</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-2085 (57)</td>
<td>-2088 (57)</td>
<td><strong>-2090 (57)</strong></td>
<td>-2089 (57)</td>
</tr>
<tr>
<td>2</td>
<td>-5774 (74)</td>
<td>-5863 (74)</td>
<td>-5835 (74)</td>
<td>-5833 (74)</td>
</tr>
<tr>
<td>3</td>
<td>-3912 (108)</td>
<td>-5922 (108)</td>
<td>-5918 (108)</td>
<td>-5914 (108)</td>
</tr>
<tr>
<td>4</td>
<td>-3638 (71)</td>
<td>-3669 (71)</td>
<td>-3667 (71)</td>
<td>-3665 (71)</td>
</tr>
<tr>
<td>5</td>
<td>-3941 (75)</td>
<td>-3967 (74)</td>
<td>-3953 (74)</td>
<td>-3955 (75)</td>
</tr>
<tr>
<td>7</td>
<td>-4943 (82)</td>
<td>-5052 (82)</td>
<td>-5013 (83)</td>
<td>-5010 (83)</td>
</tr>
<tr>
<td>8</td>
<td>-2415 (79)</td>
<td>-2427 (80)</td>
<td>-2420 (79)</td>
<td>-2419 (79)</td>
</tr>
<tr>
<td>9</td>
<td>-4211 (65)</td>
<td>-5066 (75)</td>
<td>-5054 (75)</td>
<td>-5051 (75)</td>
</tr>
<tr>
<td>10</td>
<td>-5007 (78)</td>
<td>-5059 (78)</td>
<td>-5032 (78)</td>
<td>-5031 (78)</td>
</tr>
<tr>
<td>11</td>
<td>-4316 (72)</td>
<td>-4358 (72)</td>
<td><strong>-4359 (71)</strong></td>
<td>-4358 (72)</td>
</tr>
<tr>
<td>12</td>
<td>-4964 (71)</td>
<td>-5001 (71)</td>
<td>-4983 (71)</td>
<td>-4982 (71)</td>
</tr>
<tr>
<td>13</td>
<td>-6160 (68)</td>
<td>-6210 (67)</td>
<td><strong>-6213 (68)</strong></td>
<td><strong>-6213 (68)</strong></td>
</tr>
<tr>
<td>14</td>
<td><strong>-2842 (83)</strong></td>
<td>-2841 (83)</td>
<td><strong>-2842 (83)</strong></td>
<td>-2840 (83)</td>
</tr>
<tr>
<td>15</td>
<td><strong>-4728 (92)</strong></td>
<td>-4724 (92)</td>
<td>-4724 (92)</td>
<td>-4723 (91)</td>
</tr>
<tr>
<td>16</td>
<td>-1660 (73)</td>
<td>-1664 (73)</td>
<td><strong>-1664 (73)</strong></td>
<td><strong>-1664 (73)</strong></td>
</tr>
<tr>
<td>17</td>
<td>-3413 (72)</td>
<td>-3498 (72)</td>
<td>-3458 (72)</td>
<td>-3459 (72)</td>
</tr>
<tr>
<td>18</td>
<td>-4426 (73)</td>
<td>-4478 (73)</td>
<td>-4436 (73)</td>
<td>-4097 (71)</td>
</tr>
<tr>
<td>19</td>
<td>-4450 (79)</td>
<td>-4460 (78)</td>
<td>-4454 (79)</td>
<td>-4458 (79)</td>
</tr>
<tr>
<td>20</td>
<td>-1452 (71)</td>
<td>-1461 (71)</td>
<td>-1452 (71)</td>
<td>-1453 (70)</td>
</tr>
<tr>
<td>21</td>
<td>-5603 (78)</td>
<td>-5636 (79)</td>
<td>-5628 (79)</td>
<td>-5626 (79)</td>
</tr>
<tr>
<td>22</td>
<td>-716 (57)</td>
<td>-714 (57)</td>
<td>-716 (57)</td>
<td><strong>-718 (57)</strong></td>
</tr>
<tr>
<td>23</td>
<td>-2204 (76)</td>
<td>-2248 (76)</td>
<td>-2219 (76)</td>
<td>-2218 (76)</td>
</tr>
<tr>
<td>24</td>
<td>-617 (59)</td>
<td>-627 (59)</td>
<td>-620 (59)</td>
<td>-619 (59)</td>
</tr>
<tr>
<td>25</td>
<td>-1913 (55)</td>
<td>-1937 (55)</td>
<td>-1927 (55)</td>
<td>-1927 (55)</td>
</tr>
<tr>
<td>26</td>
<td>-1934 (52)</td>
<td>-1940 (52)</td>
<td>-1939 (52)</td>
<td>-1937 (52)</td>
</tr>
<tr>
<td>27</td>
<td>-2206 (54)</td>
<td>-2207 (54)</td>
<td><strong>-2207 (54)</strong></td>
<td>-2206 (54)</td>
</tr>
<tr>
<td>28</td>
<td>-2112 (51)</td>
<td>-2114 (51)</td>
<td>-2110 (51)</td>
<td>-2109 (51)</td>
</tr>
<tr>
<td>29</td>
<td>-1402 (52)</td>
<td>-1466 (52)</td>
<td>-1436 (52)</td>
<td>-1438 (52)</td>
</tr>
</tbody>
</table>

Table 5.3: WAIC values for individual subject fits by model variant.
Figure 5.6: Estimates of the WAIC for pooled observations across all subjects for each model variant. Error bars represent two standard errors above and below each value.
have few observations on which parameter estimates can be based, leading to higher variability for these estimates. In turn, this higher variability will result in a greater penalty for the model complexity, leading to higher values of the WAIC. This increase in deviance can occur despite the good fit of the other model components and the need of the log-normal process to lessen the impact of the few outliers that are present. Furthermore, the model variants I compare here have additional terms added to account for a very small proportion of the data. It is unsurprising that the inclusion of the post-error component to models two to four led to only small improvements in the WAIC, as this led to a change in likelihood values for only a few observations.

Because the WAIC provided little information about model performance in this case, I now turn to the posterior predictive checks, which were much more informative about the effectiveness of the model.

5.5 Posterior Predictive Checks

The posterior predictive checks proved quite informative in detailing the aspects of the data that my model fit quite well or found challenging.

**The Simon effect.** Figure 5.7 shows the quantile-probability plots of correct and error responses for each congruency condition for four select subjects. These particular subjects were chosen to demonstrate the model’s ability to capture the Simon effect, some challenges it had in fitting the error response time distribution for the congruent condition, and the model’s flexibility in capturing the anomalous performance for two subjects. The ellipses indicate the 95% credible intervals for the model’s predictions for each specific subject, with red and blue denoting the congruent and incongruent conditions respectively. The green ellipses indicate the interval for the
Figure 5.7: Quantile-probability plots of correct and error responses per congruency condition for select subjects. Ellipses are the 95% credible intervals for model predictions.
model's prediction for the median response times in the incongruent condition (i.e. the condition that sequential sampling models find most challenging to fit).

Subject 22 has some of the best fits in terms of the Simon effect out of all 29 subjects, demonstrating that the model can capture the Simon effect. Subjects 3 and 16 provide a demonstration of the model’s flexibility and accounting for some abnormal response patterns. The model is able to capture the extremely long tail for the congruent error response time distribution exhibited by Subject 3, and furthermore is able to account for the at chance performance of Subject 16 in the incongruent condition. Note that for Subject 16, while the model does not quite capture the response time quantiles in the incongruent condition, it does capture the upward shift of the error response time distribution compared to the correct response time distribution (a pattern of data contrary to the pattern found with the Simon effect).

Subject 4 shows good fits for the incongruent condition, but the model fails to capture the response time distribution for errors in the congruent condition, in particular missing the lower tail of the distribution. This misfit is noteworthy as it emphasizes that the model is limited in the range of predictions that it can make for the error response time distribution in the congruent condition. Subject 4, like Subject 3, also had data inconsistent with the Simon effect. If the Simon effect was present, the error response time distribution in the congruent condition should be shifted upwards compared to the correct response time distribution. It is unusual then that the median point of Subject 4’s error response time distribution is equal to that of the median for the correct response time distribution, and the lower tail of the error distribution is in fact lower compared to the correct distribution. The model, meanwhile, can only
predict that the error response time distribution in the congruent condition is shifted upwards compared to the distribution for correct response times.

Out of all the subjects there were 8 more who exhibited posterior predictive fits for the Simon effect comparable to Subject 22, in which the model fit the data quantitatively (i.e. the credible intervals for the response time quantiles predicted by the model overlapped with the observed data). An additional seven subjects exhibited qualitative fits (in which 85 to 90% of the credible intervals for the model predictions overlapped with the observed quantiles) though the model did not fully capture the variability in the upper quantiles, in which there was less data. The remaining 13 subjects exhibit various degrees of misfit. In particular, seven subjects had error response time distributions in the congruent condition that were inconsistent with the Simon effect, due to how the 0.1, 0.3, and sometimes even the 0.5 response time quantiles for errors were shifted downward compared to the same quantiles for correct responses. Still, the model provided impressive fits to the Simon effect, considering how challenging this pattern of data can be for sequential sampling models (Servant et al., 2014).

**Trends in accuracy over blocks.** Figure 5.8 shows the posterior predictive checks of four subjects’ accuracy per block, with the model’s predictions regarding the median accuracy and its associated 95% credible intervals denoted by gray lines. The data and predictions for Subject 14, 19, 6, and 15 are shown because these subjects displayed substantial changes in accuracy over blocks and sessions. Subject 16 was the only other subject to display a similar amount of change over time. All other subjects exhibited few to no trends in their accuracy. The figure shows that the model
Figure 5.8: Accuracy per block and session for select subjects. Gray lines represent the median and lower and upper limits of the 95% credible intervals for the model predictions.
was able to account for fatigue effects in accuracy quite well (as seen in the data for Subject 14 and 19), but has trouble when there is large variability in accuracy over blocks (Subject 15) or when there is a large degree of improvement in accuracy following quite poor performance initially (Subject 6). In general though, there were few trend effects, and the model was able to account for variability in accuracy quite nicely.

**Trends in response times over blocks.** Figure 5.9 shows the posterior predictive checks of a subject’s overall response times per block, with the grey bands denoting 95% credible intervals for the model predictions of the median response time (dark gray), the 0.3 to 0.7 quantiles (medium gray), and the 0.1 and 0.9 quantiles (light gray). Superimposed over this are line segments showing the observed 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles for each block. The figure shows the data and predictions for Subject 9, 16, 14, and 20. The response time quantiles for Subject 9 became faster and the separation between the quantiles shrunk over blocks, but as a whole changes over blocks and sessions were slight, and the model accounts for this data well. The data for Subject 16 exhibit a sharper decline and a greater shrinkage in the separation between quantiles, but this pattern is still well accounted for by the model.

By contrast, Subject 14 and 20 provided challenging patterns of data in the first session that the model could not capture. The model does not capture the large degree of slowing exhibited by Subject 14 because the response time distribution for the final block had a narrow degree of separation between quantiles and a faster median compared to the preceding blocks. The model also fails to capture the large
Figure 5.9: The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles by block for select subjects. Gray bands represent the 95% credible intervals for model predictions regarding the median (dark gray), 0.3 and 0.7 (medium gray) and 0.1 and 0.9 (light gray) quantiles.
degree of heteroscedasticity exhibited in the initial four blocks for Subject 20. Note however, that fits are much better for the second session. As a whole, the model did a good job accounting for the general trends over blocks in response times.

**Post-error behavior.** To evaluate model fit to post-error behavior, I first compared the models using posterior predictive checks to double-check my conclusions based on the WAIC. I examined response time quantiles for post-error trials based on whether the preceding errors were either faster or slower compared to the previous trial. Figure 5.10 presents four quantile-probability plots, one for each model variant, showing quantiles for Subject 12’s post-error response times based on the proportion of fast versus slow errors. The 95% credible intervals for the model predictions are shown as ellipses. As can be seen in the figure, for the first model, there is underestimation of response times, which makes sense as the model has no slowing mechanism. The estimates for second model are shifted upward, now providing good fit to Subject 12’s data. The final two models show good fit to the responses following fast errors, but underestimate the response times for slow errors once again. These findings support the conclusions drawn from the WAIC. In other words, subjects appear to engage in post-error slowing regardless of the speed of the preceding error, contrary to my original hypothesis. Furthermore, note that there is little difference in predictions between the third and fourth model, suggesting that there is insufficient information to draw conclusions about trends across time for post-error behavior in the case of this particular task.

Focusing on the second model, Figure 5.11 shows the observed response time quantiles for trials following an error by each subject. Gray rectangles represent the
Figure 5.10: Quantile-probability plots for each model variant showing the 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles for Subject 12, based on whether the preceding error was fast or slow.
Figure 5.11: The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles for trials following an error per each subject. Gray boxes denotes the 95% credible intervals for model predictions, with the median represented in dark gray.
95% credible intervals for model predictions. Generally, the model does a good job accounting for the observed quantiles, though it tends to underestimate the upper tails. However, the model captures the median response times quite well. Initially at least, this suggested good fit by the model. I then examined group level effects by looking at point estimates of the model.

5.6 Associations with Perfectionism

When testing the associations between the coefficients from the cognitive model and the inventory scores, I first wanted to determine that there was sufficient variability in the inventory responses, which would indicate a good mix of subjects with a high or low degree of perfectionist characteristics. Figure 5.12 provides histograms of the summed scores for the “Doubts about actions,” “Concerns over mistakes,” and STAI subscales by session. The separation by session helps provide insight into the reliability of these measures, because most subjects completed two sessions. The test-retest reliability, as assessed by correlations, was 0.88 for the “Concerns over mistakes” subscale, 0.64 for “Doubts about actions” subscale, and 0.63 for the STAI subscale on negative affect. The values for the latter two subscales are moderate, most likely reflecting the small number of items in the “Doubts about actions” subscale, and the examinations of states for the STAI. Nonetheless, these values were high enough that, for convenience, the average summed score across session for each subscale was used in the analysis.

I first estimated the posterior modes for the relevant coefficients of the second model. These were, for the threshold, $\kappa$, $\kappa_{P1}$, $\kappa_{b1}$, $\kappa_{b2}$, and $\kappa_{M}$. In turn, the coefficients of interest for the drift rate were $\xi_2$, $\xi_{ib_1}$, $\xi_{ib_2}$, $\xi_M$, and $\xi_{S2}$. Using Stan, I then fit
Figure 5.12: Histograms of the summed scores for the “Doubts about actions”, “Concerns over mistakes”, and STAI subscales by session.
Table 5.4: A summary of my hypotheses for the point estimates of the cognitive model and their group-level effects ($\beta_0$) as well as their associations with the “Concerns over mistakes” ($\beta_{CoM}$) and “Doubts about actions” ($\beta_{DaA}$) perfectionism subscales. Hypotheses labeled as confirmed had over 95% of the posterior mass above the predicted cutoff value. Hypotheses labeled as marginal had over 90% of the posterior mass over the cutoff point.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$\kappa_{P_1} &gt; 0$ Confirmed</td>
</tr>
<tr>
<td>2</td>
<td>${\kappa_M, \xi_M} &gt; 0$ Confirmed</td>
</tr>
<tr>
<td>$\beta_{CoM}$</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$\kappa_{P_1} \approx 0$ Confirmed</td>
</tr>
<tr>
<td>4</td>
<td>${\kappa, \kappa_M} &gt; 0$ Marginal</td>
</tr>
<tr>
<td>5</td>
<td>$\xi_M &gt; 0$ Falsified</td>
</tr>
<tr>
<td>$\beta_{DaA}$</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>$\kappa &gt; 0$ Marginal</td>
</tr>
<tr>
<td>7</td>
<td>${\kappa_{P_1}, \kappa_M, \xi_M} &gt; 0$ Falsified</td>
</tr>
</tbody>
</table>

the multivariate normal model defined in equation 3.10. Following a burn in period of 250 iterations, an additional 3,000 samples were drawn for each parameter. The Geweke diagnostic value across all of the chains ranged from -0.13 to 0.10. The chains exhibited excellent convergence, and I decided no thinning was necessary. As a useful reference in the presentation of my results, I summarize my hypotheses regarding the group-level effects and associations with the perfectionism subscales in Table 5.4.

In the following figures, I display the 95%, 90%, 80%, and 50% posterior credible intervals as well as the median value for each set of regression coefficients. In this way I can examine which distributions overlap 0, which in turn allows me to draw conclusions regarding group level effects and the predictive power of the different subscales.
Figure 5.13: Credible intervals (95%, 90%, 80%, and 50%) and median estimates for the intercept term of each point estimate from the cognitive model.
5.6.1 Group-level Effects

Figure 5.13 presents the credible intervals for the intercept term for each point estimate taken from the cognitive model. Median values marked in blue indicate that the 95% credible interval does not overlap 0, while median values marked in red indicate the 90% credible interval does not overlap 0. The figure shows that the intercept term for $\kappa_{P1}$, $\xi_2$, and $\xi_{S2}$ are all substantially higher than 0. This suggests first that subjects in general exhibited increased response caution following an error, providing support for Hypothesis 1. Second, subjects had higher drift rates in the second session for both the decision and interference process (most likely reflecting a general increase in processing speed due to practice). The intercept for $\kappa$ is also greater than 0, though this is less interesting as it is extremely unlikely for the threshold to be near 0 (since this would imply subjects were making nearly instantaneous responses).

In turn, $\kappa_{b1}$ was substantially less than 0, suggesting a gradual decrease in response caution over blocks for the first session. A similar, though non-significant, trend is evident for $\kappa_{b2}$, as over 90% of the posterior mass is distributed below 0. The 90% credible intervals for both $\kappa_M$ and $\xi_M$ lie above 0, suggesting that there is a small but reliable increase in response caution and drift rates during the motivation condition, providing support for Hypothesis 2. Recall that for my exploratory analyses in section 5.2, there was no noticeable impact of payoffs on response times. The finding of slightly higher response caution and drift rates based on payoffs emphasizes how sequential sampling models can increase the power to detect small effects.
Figure 5.14: Credible intervals (95%, 90%, 80%, and 50%) and median estimates for the “Concerns over mistakes” term of each point estimate from the cognitive model.
5.6.2 Associations with “Concerns over Mistakes”

Figure 5.14 presents the credible intervals for the regression term describing the association between the “Concerns over mistakes” subscale and point estimates from the cognitive model. In a similar vein to previous studies (Schrijvers et al., 2010; Tops et al., 2013), the “Concerns over mistakes” subscale did not significantly predict any point estimate. This confirms Hypothesis 3, in that I expected no relation between “Concerns over mistakes” summed scores and $\kappa_{P1}$, the increase in response caution following an error. I found marginal support for Hypothesis 4, in which higher scores on the “Concerns over mistakes” subscale would predict an increase in the coefficients $\kappa$ and $\kappa_M$. Although the 95% posterior credible intervals contained 0 for both coefficients, the 90% posterior credible interval was higher than 0. I did not find any support for Hypothesis 5, in which higher scores on the “Concerns over mistakes” subscale would predict an increase in drift rates.

5.6.3 Associations with “Doubts about Actions”

Figure 5.15 presents the credible intervals for the regression term describing the association between the “Doubts about actions” subscale and point estimates from the cognitive model. The subscale showed a signification negative association with $\xi_M$, indicating that greater doubts about actions corresponded to poorer evidence sampling in the payoffs condition. This association is in the opposite direction of what I expected based on Hypothesis 7. Additionally, contrary to Hypothesis 7, there was no association between the subscale scores and $\kappa_M$ and $\xi_M$. There was, however, marginal support for Hypothesis 6, as the 90% posterior credible interval for $\kappa$ lay above 0. This result suggests that subjects with more doubts about their actions
Figure 5.15: Credible intervals (95%, 90%, 80%, and 50%) and median estimates for the “Doubts about actions” term of each point estimate from the cognitive model.
were marginally more cautious. Furthermore, the 90% posterior credible intervals for \( \xi_2 \) and \( \xi_{b_1} \) lay below 0.

### 5.6.4 Associations with the STAI Subscale

Figure 5.16 presents the credible intervals for the regression term describing the association between the STAI subscale on negative affect and point estimates from the cognitive model. Most of the posterior distributions overlapped substantially with 0, but the 90% credible interval for \( \xi_{S_2} \) fell below 0. This suggests that subjects who on average indicated more current feelings of negative affect made slower responses from the interference process in the second session.

### 5.6.5 Correlations between Coefficients

By fitting the multivariate normal model defined in equation 3.10, I also obtained correlations between the point estimates for the coefficients (the matrix \( \Omega \)) and the variances of the coefficients (the vector \( \eta \)). I report in Table 5.5 the correlations in the off-diagonal cells and the variances in the diagonal cells. I denote with an asterisk values whose 95% credible intervals did not include 0.

The coefficient \( \kappa_2 \) had a moderate negative correlation with and \( \xi_2 \) and a positive correlation with \( \xi_{b_2} \). This suggests that subjects who showed an increase in response caution for the second session also had a lower drift rate, but exhibited a larger increase in drift rate over blocks. The coefficient \( \xi_2 \) also had a moderate positive correlation with \( \xi_{b_1} \) and \( \xi_{S_2} \), and a negative correlation with \( \xi_{b_2} \). This suggests subjects that showed an increase in drift rate for the second session also had a gradual increase in drift rate over blocks for the first session, but a gradual decrease in drift rate for the second block. Higher drift rates in the second session for the main decision process
Figure 5.16: Credible intervals (95%, 90%, 80%, and 50%) and median estimates for the STAI term of each point estimate from the cognitive model.
also corresponded with higher drift rates for the interference process. The coefficient \( \xi_{b1} \) had a weak negative correlation with \( \xi_{b2} \), suggesting again that subjects with a gradual improvement in drift rate for the blocks in the first session had a gradual decline for the second session. Additionally, \( \kappa_M \) and \( \xi_M \) were moderately positively correlated. Subjects who became more cautious in the motivation condition also had improved drift rates. As a whole, parameter coefficients correlated in an intuitive and reasonable manner, and the significant (and often sizeable) correlations emphasize that such inter-relations cannot be ignored when evaluating the group level effects.

5.7 Further Analyses of Post-error Adjustments

The group level analyses of the model estimates suggested that subjects exhibited a significant increase in response caution following an error. If a person was indeed becoming more cautious, then the model predicts a corresponding increase in accuracy following an error. This emphasizes one of the strengths of sequential sampling models and quantitative models in general, their ability to generate strong and testable hypotheses. Therefore, I further examined the post-error behavior of subjects and the model predictions by looking at the correct and error response time distributions for responses that were preceded by an error. Figure 5.17 details quantile-probability plots of correct and error response time distributions for trials following an error for Subject 17, 27, 2, and 8. The model fits the pattern of data provided by Subject 17 quite well. Subject 2 shows slight misfit in the upper tail of the response time distribution which is underestimated.
Table 5.5: Correlations between point estimates for coefficients from the cognitive model, with variances on the diagonal. Asterisks indicate that the 95% credible intervals for the correlations do not overlap 0.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>$\kappa$</th>
<th>$\kappa_{P_1}$</th>
<th>$\kappa_{b_1}$</th>
<th>$\kappa_{b_2}$</th>
<th>$\kappa_M$</th>
<th>$\xi_2$</th>
<th>$\xi_{b_1}$</th>
<th>$\xi_{b_2}$</th>
<th>$\xi_M$</th>
<th>$\xi_S_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\kappa_{P_1}$</td>
<td>-0.12</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\kappa_{b_1}$</td>
<td>-0.15</td>
<td>0.03</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\kappa_{b_2}$</td>
<td>-0.18</td>
<td>0.12</td>
<td>0.09</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\kappa_M$</td>
<td>0.15</td>
<td>-0.06</td>
<td>-0.01</td>
<td>0.1</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\xi_2$</td>
<td>-0.07</td>
<td>0.11</td>
<td>-0.16</td>
<td>-0.59*</td>
<td>-0.26</td>
<td>0.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\xi_{b_1}$</td>
<td>0.01</td>
<td>0.03</td>
<td>0.2</td>
<td>-0.28</td>
<td>0.06</td>
<td>0.48*</td>
<td>1.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\xi_{b_2}$</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.05</td>
<td>0.68*</td>
<td>0.24</td>
<td>-0.64*</td>
<td>-0.34*</td>
<td>1.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\xi_M$</td>
<td>-0.1</td>
<td>0.03</td>
<td>0.19</td>
<td>0.12</td>
<td>0.4*</td>
<td>0.13</td>
<td>0.06</td>
<td>-0.07</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>$\xi_S_2$</td>
<td>-0.02</td>
<td>0.02</td>
<td>-0.24</td>
<td>-0.25</td>
<td>-0.12</td>
<td>0.32*</td>
<td>-0.11</td>
<td>-0.14</td>
<td>0.04</td>
<td>0.47</td>
</tr>
</tbody>
</table>
Figure 5.17: Quantile-probability plots of correct and error response times for trials following an error for select subjects. Ellipses are the 95\% credible intervals for model predictions.
Subject 27 and 8 reveal an important falsification of the post-error mechanisms in the current model. In both cases, while the response time quantiles are generally well accounted for, the model substantially overestimates the accuracy of the subjects following errors. This overestimation occurs for a further 16 subjects. This falsification of the post-error mechanism for the model provides extremely useful information, as it indicates that, for over half of the sample, the post-error slowing exhibited by subjects cannot be attributed to increased response caution. This finding has several important implications for the use of interference tasks and measures of post-error slowing in the examination of perfectionism, which I will elaborate upon in the following chapter.

5.8 Summary

An examination of aggregate measures of performance and their standard errors suggests that subjects reliably produced the Simon effect and exhibited post-error slowing. In contrast, there appeared to be no fatigue effects, only a mild practice effect, and no effect of payoffs on response time and accuracy. A main outcome of these analyses is that aggregate measures may mask important relations in the data. These findings provided a useful baseline to compare against my model fit.

I used the WAIC to compare my four model variants: a version without any post-error behavior, a version with an increase in response caution following errors, a version with an increase in response caution following fast errors only, and a variant of the third model that had separate coefficients for the first and last half of a session. While the standard errors overlapped for all four models, the second model had the lowest deviance criterion overall, suggesting that while mechanisms for post-error
slowing were necessary, there is no evidence to suggest that subjects engaged in multiple types of post-error adjustments.

Posterior predictive checks gave more information about the performance of the model. The model provided a good account of the Simon effect that was present in the data of 16 subjects, capturing in the incongruent condition the downward shift of the error response time distribution compared to correct choices, and in congruent condition the upward shift of the error response time distribution for errors compared to the correct responses. The model displayed flexibility in accounting for some of the abnormal response patterns of subjects, including the ability to capture extremely long upper tails in response time distributions and surprisingly poor performance. Seven of the remaining 13 subjects displayed an abnormal pattern in the congruent condition (which was masked when examining the aggregate measures of performance) in which the 0.1, 0.3, and sometimes even the 0.5 response time quantiles for errors were in fact faster relative to the correct response time quantiles. This pattern of data is inconsistent with the Simon effect, and the model was unable to account for such findings.

In the few cases in which subjects exhibited substantial trends in accuracy and response time over blocks, the model was able to capture fatigue effects in accuracy and practice effects in response times, but struggled with large improvements in accuracy over blocks and extreme variability in accuracy and response times. Most subjects did not exhibit a large degree of trend in their data, and unsurprisingly the model did quite well in accounting for this behavior.

Initial explorations of post-error behavior found, first, further support that while subjects did engage in post-error slowing, there was no need to account for a lack
of slowing following slower errors. Second, quantile-probability plots of post-error response times conditioned on the speed of the error found that the model with a single mechanism for post-error adjustments, where subjects increased their thresholds following an error, had the best fit. Third, an examination of response time quantiles suggested that, while the model underestimated the upper tail of the distribution (in which there are very little data), it did a good job in capturing the response time distributions for trials following an error.

Examinations of point estimates of the cognitive model at the group level using a multivariate normal model revealed several interesting effects and associations (or lack thereof) with inventory scores. First, there was a reliable increase in response caution following errors across the sample. There was also a small but reliable increase in both response caution and drift rates for the motivation condition, suggesting this manipulation was more successful than aggregate measures of performance would have led us to believe. Summed scores for the “Concern over mistakes” subscale did not significantly predict any point estimates, which was expected for the post-error shift. Though non-significant, 90% of the posterior mass for the threshold and its shift based on payoffs was above 0, which was in the predicted direction.

Summed scores for the “Doubts about actions” subscale predicted a drop in drift rates for the payoffs condition, a finding that was in the opposite direction to that of Hypothesis 7. As was the case with the “Concern over mistakes” subscale, the posterior distribution for the coefficient $\kappa$ once again had 90% of its mass above 0, though as before this was non-significant. The summed scores for the STAI subscale on negative affect had no associations with point estimates except for a decrease in the shift in sessions for the drift rate of the interference process.
The group-level analyses indicated that subjects exhibited a reliable increase in response caution following errors, as evidenced by point estimates of $\kappa_{P1}$ being consistently higher than 0. An important consequence of this finding is that it produces a directly testable hypothesis: if subjects truly are becoming more cautious following errors, they should become more accurate in post-error trials. I tested this hypothesis by examining the posterior predictive checks for the correct and error response time distributions on trials following an error. The model substantially overestimated the accuracy of 18 subjects, indicating that increased response caution is not the mechanism that best accounts for the post-error slowing seen in most of the data.
Chapter 6: Discussion

Perfectionism, typically conceptualized as the setting of high standards of performance accompanied by overly critical evaluations of one’s own behavior (Frost et al., 1990), has been implicated in the cause, maintenance, and course of several psychopathological states (Shafran et al., 2002; Shafran & Mansell, 2001). Despite the growing research on perfectionism (Flett & Hewitt, 2002), it remains an ill-defined and poorly understood construct (Egan et al., 2011).

One solution to this problem is to develop a better understanding of the basic cognitive processes contributing to perfectionism. Recent work has tried to address this by examining relations between perfectionism and performance monitoring in simple choice tasks (Pieters et al., 2007; Schrijvers et al., 2010; Tops et al., 2013). The studies by Schrijvers et al. (2010) and Tops et al. (2013) found that the ERN, believed to represent activity in the ACC for detecting errors and other unexpected events, correlated with subjects’ degree of doubts about actions, while the Pe, believed to represent an emotional response and conscious awareness of errors, correlated with subjects’ degree of concern over mistakes.

However, the subscales for perfectionism showed little relation with behavioral measures, in particular post-error slowing, which typically is thought to be the behavioral adjustment from recruiting greater cognitive control to avoid mistakes. This
is a surprising finding, as the hypothesis of greater post-error slowing corresponding
with a higher degree of perfectionism is intuitive and appealing. The issue is com-
plicated by the fact that performance monitoring and post-error behavior is affected
by motivation (Pailing & Segalowitz, 2004), as well as by findings of global changes
over the course of an experiment in both post-error slowing and the ERN (Luu et
al., 2000). Furthermore, post-error slowing is a complex, sensitive, and poorly under-
stood measure. Much of the research applying post-error slowing to study individual
differences relies on aggregate measures that can mask important effects and lead
to spurious findings (Dutilh, van Ravenzwaaij, et al., 2012) especially if motivation
changes or a subject becomes fatigued during a session. Finally, the studies by Pieters
et al. (2007), Schrijvers et al. (2010), and (Tops et al., 2013) did not test their as-
sumption that post-error slowing represented a more cautious mode of responding,
despite research that has found that post-error slowing can in fact reflect less cautious
modes of responding (e.g. Notebaert et al., 2009).

It would be pragmatic and useful to be able to rely on behavioral measures and
still have the ability to draw conclusions about the underlying cognitive mechanisms
for perfectionism. The lack of associations between inventory scores of perfectionism
and post-error slowing is therefore problematic. However, it is clear that a more ro-

dust methodology is needed to address the potential confounds undermining the use
of this measure. Confounds like global trends in response times, fatigue and prac-
tice, as well as a mix of post-error adjustments all may have masked associations
between post-error slowing and measures of perfectionism. Therefore, I relied on a
quantitative model of simple choice and response time to examine underlying cogni-
tive mechanisms with greater precision. I then used the estimates from this model
to examine associations with a measure of perfectionism, the FMPS (Frost et al., 1990). I was able to better control for confounds, and critically, I explicitly tested the assumption that post-error slowing represented an increase in caution.

I developed a novel model for the Simon task to test my hypotheses. I predicted that after controlling for global trends, motivation, and for a combination of increased cautiousness following fast errors, but none following slow responses, that I would find significant associations between the shift in thresholds following errors and the subscales of the FMPS. Specifically, I predicted that subjects with higher summed scores for the “Doubts about actions” subscale would show higher shifts in thresholds following an error.

Based on the work of Luu et al. (2000) and Tops et al. (2013), I also predicted that subjects with higher scores on the “Concern over mistakes” subscale would exhibit a decline in post-error slowing over blocks. An easy way to examine this is to have two separate shifts in thresholds following an error, one for the first half of a session, and one for the second half. If a subject engages in less post-error slowing over blocks, then the first threshold shift should be higher relative to the second, and I predicted that the difference between the two shifts would correlate with the “Concern over mistakes” subscale. I found instead, that for 18 of my 29 subjects, the significant degree of post-error slowing had nothing to do with increased cautiousness, thereby falsifying one of the core but untested assumptions made by the previous studies.

6.1 Implications

I now discuss the implications my results have for future research, consider some useful methodological extensions, and end with some brief conclusions.
6.1.1 Post-Error Slowing

While subjects did engage in a significant degree of post-error slowing, my analyses demonstrated that such slowing could not be attributed to a more cautious mode of responding. My model accounted for the post-error slowing by an increase in thresholds, and did well in accounting for the post-error response time distributions. However, posterior predictive checks found that the model substantially overestimated post-error accuracy for 18 subjects. Subjects in fact became less accurate following an error. Hence, post-error slowing was not attributable to increased response caution, making the lack of a correlation between post-error slowing and the perfectionism subscales much less surprising. Furthermore, this finding raises many new and interesting questions.

For instance, if subjects are not becoming more cautious, then what is driving the post-error slowing? I propose two possibilities based on past literature. First, as noted in my introduction, recent work has shown that an orienting account instead of an action monitoring account can better explain post-error slowing under certain tasks and situations (Notebaert et al., 2009; Núñez Castellar et al., 2010; Wessel et al., 2012). This theory posits that post-error slowing is due to unexpected events orienting attention away from the task. Note that in my task, incongruent trials occurred infrequently and had the highest error rates. The infrequent occurrence of such trials could contribute to them being more surprising compared to congruent trials. As noted by Dutilh, Vandekerckhove, et al. (2012), for sequential sampling models, the orienting account of post-error slowing leads to the prediction that subjects should have reduced drift rates following errors, reflecting the interfering effect of surprise.
over the previous outcome. The reduced drift rate predicts slower, less accurate responses.

Subjects may also become biased against the choice made in error (Laming, 1968, 1979; Rabbitt & Rodgers, 1977). This hypothesis has some additional support based on the finding that following an incongruent trial, subjects are much less likely to display the Simon effect (e.g. Proctor et al., 2013). Such a possibility can be tested by allowing unequal thresholds, with a higher threshold for the choice that was previously selected in error. An additional consideration, tying into the musings of Rabbitt (1966) and my own concerns presented in my introduction, is that subjects engage in all three of types of the post-error adjustments discussed here. I have shown that it is possible to test for distinct types of error adjustments through careful specification of covariates, so such a possibility should not be ignored.

6.1.2 Perfectionism and ERPs

My results, while providing an explanation for the lack of any associations between inventory scores for perfectionism and post-error slowing, also suggest that the conclusions drawn by previous studies regarding perfectionism and the ERN and Pe components should be interpreted more cautiously.

Research into post-error slowing has always been tightly linked with EEG studies. The conflict monitoring account of post-error slowing, in which the slowing is attributed to increased response caution and greater recruitment of cognitive control, arose from work focused on understanding the mechanisms driving the ERN component (e.g. Botvinick et al., 2001, 2004). Moreover, several studies have found that post-error slowing is correlated with the amplitude of the ERN (e.g. Danielmeier et
al., 2011; Debener et al., 2005; Garavan et al., 2002; Gehring et al., 1993; Hester et al., 2007). As noted before, the account provided by conflict theory for both the ERN and post-error slowing are attractive candidates for potential cognitive mechanisms associated with perfectionist characteristics. The studies by Pieters et al. (2007), Schrijvers et al. (2010), and Tops et al. (2013) all operated under the assumption that the ERN component, the Pe component, and post-error slowing were all reflective of some form of performance monitoring. In other words, these studies assumed a conflict monitoring approach was an appropriate framework in which to interpret their data.

However, my results are not consistent with the conflict monitoring theory, due to the decrease in accuracy following errors.

An alternative, the orienting account, suggests that the past studies examining perfectionism and the ERN and Pe perhaps should have instead interpreted the behavioral and neural markers of errors and post-error responses based on mechanisms for surprise. For instance, the EEG study by Núñez Castellar et al. (2010) found that slowing following infrequent errors (or infrequent correct responses) was not predicted by the ERN component. Instead, the slowing was significantly predicted by the P3 component, a component that among other things, reflects attentional processes sensitive to novel events. It may be less appropriate to focus on the ERN and Pe components for data more consistent with an orienting account.

An important question is whether my findings in the context of the Simon task can generalize to the studies by Pieters et al. (2007), Schrijvers et al. (2010), and Tops et al. (2013) that relied on the Ericksen flanker task. While both designs are interference tasks, this does not necessarily mean the errors and post-error behaviors will be
similar across tasks. However, Tops et al. calculated a measure of post-error accuracy, subtracting the proportion of accurate responses following an correct responses from the proportion of accurate responses following an error. The authors report that the average of this measure was 0.03 with a standard deviation of 0.07. In other words, accuracy following an error was not necessarily higher than the accuracy following a correct response, contrary to the predictions of the conflict monitoring theory (Note that Schrijvers et al. appears to have not analyzed post-error accuracy at all).

Furthermore, a study by Riba, Rodríguez-Fornells, Münte, and Barbanoj (2005) demonstrated a dissociation between the ERN component and post-error slowing using a flanker task. The authors found that administering a benzodiazepine known as alprazolam to subjects led to a decreased amplitude for the ERN component, but had no effect on post-error slowing. Therefore, it is possible that post-error slowing in the flanker task, just I found for the Simon task, will not always represent increased cognitive control. Future researchers relying on interference tasks to examine the cognitive mechanisms of perfectionism should consider additionally focusing on other ERPs, such as the P3 component, which could lead to better insights in the links between the behavioral and neural results. Alternatively, researchers could consider different experimental designs more likely to produce post-error behavior based on increased response caution.

6.1.3 Adaptive Individual Deadlines

My experimental design differed in another important way compared to past studies examining the cognitive mechanisms of perfectionism. The studies by Pieters et al. (2007), Schrijvers et al. (2010), and (Tops et al., 2013) all relied on an adaptive
individual deadline to produce high error rates in their task. Subjects had to respond within a time limit (their mean response time from the previous block plus half of a standard deviation) to avoid negative feedback. The authors did not discuss whether such a deadline would have an influence on the response times for post-error response times (or, for that manner, the ERPs they measured). Sequential sampling models, however, allow a great deal of insight into how such deadlines can impact response times and choice. The imposition of a response time deadline affects the threshold (e.g., Van Zandt & Maldonado-Molina, 2004). To meet the speed emphasis, subjects must set thresholds lower, producing faster but less accurate responses. In the case of the adaptive deadline, subjects must then change their thresholds each block. However, I found this design problematic. The assumption that post-error slowing represents a more cautious mode of responding corresponds to an increase in thresholds following an error. When analyzing post-error slowing under this assumption, we are, in essence, analyzing changes in threshold.

Unfortunately, the use of the adaptive individual deadline could potentially introduce a great deal of additional systematic variability in the threshold, thereby confounding the analysis of post-error slowing. Though I chose to avoid this confound by not using an adaptive individual deadline, a beneficial extension would be to replicate the previous studies and explicitly model changes in threshold due to the deadline. In this way, we would gain insight in how such deadlines impact post-error behavior, determining whether they truly are a confound. Furthermore, such an extension could further test whether the post-error slowing found in tasks with adaptive individual deadlines is due to response caution, an assumption my project has cast into doubt.

168
6.1.4 The Simon Task

An interesting byproduct of my modeling efforts to examine perfectionism was the development of a novel model that accounts for the Simon effect. As I noted in Section 3.1.1, the Simon effect produces patterns of response times and accuracy that are extremely challenging for sequential sampling models to fit (Servant et al., 2014). The tendency of the response time distribution to shift downwards for error responses relative to the response time distribution for correct responses poses a problem, but this is exacerbated by the switch in the congruent condition in which the response time distribution for errors is shifted upward relative to the response time distribution for correct responses. Nonetheless, I was still able to obtain good fits via my mixture model. However, my project is not the first to fit sequential sampling models to data from the Simon task. I now discuss these alternative models, contrast them to my mixture model, and comment on some of the specific advantages my model possesses. I conclude with a brief discussion of some intriguing response patterns provided by a small subset of my subjects that were contrary to the patterns predicted by the Simon effect.

Previous models for interference tasks. There are two variants of the diffusion model that have been previously fit to the Simon task (Servant et al., 2014). The first variant was proposed by Hübner, Steinhauser, and Lehle (2010). The authors extended the diffusion model and proposed a dual-stage two-phase (DSTP) model of selective attention. In turn, White, Ratcliff, and Starns (2011) made a variant of the diffusion model that they called a shrinking-spotlight model (SSP). Both models involve an increase in attentional selectivity for the relevant stimulus attribute over
the course of a trial. To provide some background context, in the diffusion model (Ratcliff & McKoon, 2008) evidence accumulates towards one of a pair of thresholds. The evidence can be thought of the difference between sensory inputs supporting the separate choices. This difference is a random variable that follows a normal distribution with a mean equal to the drift rate, and a variance that, for model identification purposes, is typically fixed to either 0.1 or 1. The drift rate is assumed to remain constant throughout the time-course of a trial. The variants of the diffusion model proposed by Hübner et al. (2010) and White et al. (2011) relax this latter assumption.

**The DTSP model.** Hübner et al. (2010) suggest that during an early selection phase, the drift rate consists of the sum of two separate components, a rate for the relevant stimulus attribute (e.g. color) and a rate for the irrelevant stimulus attribute (e.g. spatial location). In incongruent trials, the rate for the irrelevant stimulus is negative. This latter aspect captures the interference caused by the irrelevant stimulus in the incongruent condition, because the rate for the irrelevant stimulus drives the net drift rate down, resulting in the process terminating at the incorrect threshold more often, producing fast errors. In turn, this mechanism can account for the fast correct responses in the congruent condition, as the rate for the irrelevant stimulus in fact improves the net drift rate.

In this early stage, the process races against a second diffusion process, a target identification process. If the latter process wins the race, the subject enters a late-stage selection, in which the composite drift rate defined above changes to a new drift rate determined solely based on the relevant stimulus attribute. Because the
irrelevant stimulus no longer plays a role in this later stage, there is an increase in the drift rate, which in turn accounts for the improved accuracy seen with slower responses. However, if the target identification reaches the incorrect boundary, this results in a slow perceptual error instead. The authors demonstrated the viability of their model by fitting it to data from the Eriksen flanker task.

**The SSP model.** In contrast, the variant of the diffusion model proposed by White et al. (2011) has the drift rate vary continuously instead of discretely over the time-course of a trial. Furthermore, while the model by Hübner et al. was intended as a general framework for interference tasks, the SSP model was designed specifically to account for the spatial dynamics in the Eriksen flanker task. The model assumes that early in a trial, a subject’s spatial attention is broad, and as such can be influenced by the flankers. As the trial progresses, a subject’s spatial attention narrows towards the central target. The model captures this by allowing drift rate at every time step to be based on the perceptual input of the target stimulus and the flanker stimuli, weighted by an allocation of attention.

It is assumed that for the standard Eriksen flanker task, the target and flanker inputs have the same quantity of evidence. The allocation of attention attributed to the target is modeled as the integral over -0.5 to 0.5 (the region of space corresponding to the flankers) of a normal distribution with a mean of 0 and a standard deviation $\sigma_a$. The variable $\sigma_a$ decreases at a linear rate $r_d$. The authors then note that for congruent trials, the drift rate is constant (the attentional weights always sum to 1). However, in incongruent trials, the drift rate is initially directed towards the
incorrect threshold, producing fast errors, but over time is driven back towards the correct threshold.

**DTSP and SSP model fits to the Simon task.** Servant et al. (2014) fit the DTSP and SSP models to a version of the Simon task in which the chroma of a colored circle was altered to vary difficulty. Half of the trials were congruent, half were incongruent. The models were fit by averaging the response time quantiles across 12 subjects, and then applying a SIMPLEX routine to minimize the \( G^2 \) likelihood ratio statistic calculated from the difference between the observed quantiles and predictions from model simulations.

The authors found that both models performed poorly. One way that the authors examined fit was by ranking the data by the speed of the response. The ranked data were then split into seven ordered bins, and for each bin the mean response time and corresponding accuracy was calculated. The Simon task produced a cross-over in mean response times and accuracy between the congruent and incongruent conditions, where for higher mean response times, the accuracy for the incongruent condition was higher relative to the congruent condition. Servant et al. note the main source of misfit for both models was their inability to account for this cross-over. The authors note that this led to collateral distortions, with the DSTP model predicting a Simon effect on mean response time that was too small and errors that are too fast in the congruent condition, while the SSP predicted errors that were too fast in all conditions. These findings emphasize how difficult the Simon task can be to fit.

**Comparisons and contrasts between the models.** There are some interesting contrasts between my mixture model and the DSTP and SSP models. The SSP model
involves a single decision process whose inputs driving the evidence accumulation dynamically change over a trial due to changes in attention allocation. In other words, the state of the system continuously evolves. The DSTP model also involves a single decision process, but the manner in which the evidence accumulation changes is different. The DSTP proposes a parallel process that modulates the original decision process resulting in a discrete shift in the state space. The target identification process can then be thought of as a discrete approximation for a mechanism governing whether a subject successfully narrows his or her attention to the target and eliminates the interfering influence of the irrelevant attribute.

Both models appear to account for the Simon effect by proposing a single decision process modulated by attention. As noted by Proctor et al. (2013), popular accounts of the Simon effect propose dual routes, a direct automatic route govern by spatial location and an indirect, more controlled route governed by the relevant stimulus attribute. These dual routes are present in the DSTP and SSP models via the inputs driving the drift rate. Both models propose an early stage in which drift rate is a composite of rates based on relevant and irrelevant attributes; these rates can be thought of as coming from the indirect and direct routes.

In contrast, the architecture of the mixture model that I developed posits two separate decision processes. On a given trial, a subject can attend to the relevant attributes of a stimulus, and their behavior will be a product of the main decision process alone. By contrast, a subject could instead attend to the irrelevant attributes of a stimulus, and in this case their behavior would be the product of the interference process alone. The probabilities governing the mixture components can therefore
be interpreted as an attentional component, representing the proportion of trials in which a subject attends or fails to attend to the relevant attribute.

The interference that the processes can exert on one another are captured by changes in the drift rates for each process based on congruency. However, in contrast to the DSTP and SSP models, I did not impose a precise mechanism governing these interactions; instead, I simply let parameters freely vary. In other words, in my model the dual routes are represented not just in the inputs, but by separate decision processes entirely. It would be beneficial if a precise mechanism governing the interaction between the processes in my model could be stipulated. Also, a comparison between the architecture for the DSTP, SSP, and my mixture model would be useful, as this comparison could allow insights into the neural mechanisms driving the interference effect, leading to hypotheses that could be evaluated using neurophysiological methods.

A key strength of the DSTP and SSP models is the precise cognitive mechanisms that they propose to account for the interference effect. However, this precision comes with the cost of complexity. Neither model has a tractable likelihood, and predictions must be based from simulations instead. This limits the ease of applicability of these models, especially in terms of fitting the models using maximum likelihood or Bayesian methods. A major strength of the model I have developed is that it has a tractable likelihood, and I therefore was able to fit the model using full non-approximate Bayesian methods. Furthermore, my model still has interpretable parameters corresponding to specific cognitive mechanisms.
Inconsistencies with the Simon effect. There is a source of misfit in my model with interesting implications regarding the typical patterns found in the Simon task. A subset of my subjects had 0.1, 0.3, and sometimes even 0.5 response time quantiles for errors in the congruent condition that were shifted downwards compared to the quantiles for correct responses. As noted before, this pattern of data is inconsistent with the Simon effect (e.g. Servant et al., 2014). These findings are particular puzzling because the subset of subjects performed well, and did not show any abnormal trends in their data. However, error responses in the congruent condition occurred the least often, making this distribution more noisy and highly susceptible to outliers. Indeed, the widest credible intervals for model predictions regarding response times occur in the error conditions, in particular the congruent condition. It is difficult to ascertain if there is a systematic effect that produces these faster tails, or if this is simply noise contaminating the data.

One possible solution to deal with the misfit is to introduce parameter variability in the residual latencies. Ratcliff and Tuerlinckx (2002) note that including variability in residual latencies led to corrections of misfits for the diffusion model corresponding to the fastest response time quantiles in data sets possessing large variability in the lower tail end of the distributions. My data exhibited similar variability, as some subjects had several extremely rapid responses. Variability in the residual latencies is a potential way to correct this. Another possibility is to introduce a fourth mixture component representing fast guesses, such as an exponential process (e.g. Craigmile et al., 2010), which would have the added benefit of removing the need to delete blocks with extremely rapid responses. Both of these are viable extensions of the model for improved fit.
6.2 Methodological Extensions

I now briefly discuss some useful methodological extensions for future work. While such extensions would be computationally or pragmatically challenging to implement, they are still worth noting.

6.2.1 Hierarchical Models

My model coefficients were estimated with uncertainty, quantified by the posterior distributions. However, when examining the group level differences using point estimates of the coefficients, it was assumed that these point estimates were measured without error. Ideally, we would like to incorporate the uncertainty around the coefficients estimates into the group level analysis. A useful extension to my project would be the use of a hierarchical model (Gelman & Hill, 2007). As noted by Gelman and Hill, hierarchical models assume that subjects are a randomly-drawn sample from a population for which certain characteristics are known. Such models allow sets of observations from individuals, conditions, or other blocking variables to each have their own set of parameters. Like the observations themselves, the parameters are treated as randomly drawn, exchangeable samples from a larger population, meaning that they can be treated as random effects in a statistical model. Individual differences are therefore incorporated into the model, since subjects each have their own set of parameters. However, group-level effects can still be examined by looking at the higher level distributions that describe the parameters for each individual (Vandekerckhove et al., 2011).

Hierarchical models possess several methodological advantages over standard models. For instance, hierarchical models are well suited to handle data sets with few
trials per subject (often the case with clinical studies), even when an individual subject may provide insufficient trials to accurately estimate all model parameters (Vandekerckhove et al., 2011). Another advantage over standard models is that hierarchical models allow more control in modeling the variability in data, thereby avoiding the negative effects of aggregation, which can be particularly severe for nonlinear models like sequential sampling models (Rouder & Lu, 2005). Finally, hierarchical models allow a great deal of flexibility in testing hypotheses (Vandekerckhove et al., 2011; De Boeck & Wilson, 2004). While the implementation of a hierarchical version of my model would be challenging from a computational perspective, nonetheless this would be a useful extension.

6.2.2 Estimation Routines

For my project, the DE-MCMC algorithm proved extremely useful in allowing me to obtain posterior estimates of my coefficients despite the high correlations between parameters that typically plague sequential sampling models (Turner et al., 2013). One aspect that contributed to the success of the DE-MCMC algorithm was the fact that I ran several hundred chains in parallel. However, there was always a small proportion of chains that performed poorly, requiring careful inspection of all the chains for each coefficient, which was a time-consuming process. This problem could perhaps be improved via more careful choices of the tuning parameters, but another option is to consider alternative samplers, such as the HMC sampler (Neal, 2011). While the implementation of such a sampler would require determining the partial derivatives for all the parameters in the model, taking the time to do so could
potentially lead to sizeable improvements in sampling from the posterior, and as such is a worthwhile consideration.

### 6.2.3 Measurement of Perfectionism

Despite the use of my cognitive model to parse variability in my data based on psychologically meaningful parameters, there were few significant associations between point estimates and the summed scores for the inventory subscales. However, for the “Concern over mistakes” subscale, slightly above 90% of the posterior mass was greater than 0 for the regression coefficients corresponding to the point estimates of $\kappa$ and $\kappa_m$. In the case of the “Doubts about actions” subscale, the regression coefficient corresponding to the point estimate of $\kappa$ had slightly above 90% of the posterior mass greater than 0, while about 90% of the posterior mass for $\xi_2$ and $\xi_{b_1}$ were below 0. This warrants further investigation, especially because theoretically, one would expect a relation between the thresholds and perfectionist characteristics. Moreover, one would also expect that perfectionistic individuals would be more greatly affected by the motivation manipulation.

One possibility that can account for the smaller effects seen in my study is the fact that the inventories measure their constructs with error. This is particularly obvious with the “Doubts about actions” subscale, which, despite supposedly representing a stable construct, had a test-retest correlation of only 0.64. Furthermore, the summed scores were used with the assumption that they were independent, which naturally is also not true, as the FMPS scales show moderate correlations with each other (Frost et al., 1990).
A useful extension for later work would be to explicitly account for the measurement error and covariances for the inventories in some way. A latent trait model taken from Item Response Theory (IRT), the Graded Response Model (GRM; Samejima, 1969) is particularly appropriate to address this problem. The GRM is a latent variable model intended specifically for polytomous responses like the 5-point Likert response set used in the FMPS and the 4-point likert-type scale used in the STAI. The model gives the probability that, given \( m \) categories, an individual will pick response category \( c \), where \( c \in \{1, 2, ..., m\} \), given his or her level on a latent construct (e.g. perfectionist tendencies) and the properties of the item. A person’s level on the latent trait can be estimated, and that estimate can be used in lieu of the summed scores. Controlling for individual item properties may allow a better representation of a person’s level on the latent construct compared to what was provided by the summed scores. To capture the covariance structure of the inventories, a multidimensional version of the GRM can be used (e.g. Wirth & Edwards, 2007), in which separate latent variables are specified for each subscale, but the covariance structure of the latent variables is explicitly estimated.

One limitation in using the GRM however, is that to get estimates for the model parameters, one typically needs a sample size in the hundreds if not thousands. A workaround that still makes this approach viable for future research would be to have a large sample of subjects from the relevant population (e.g. college students) complete the inventories, and then use their data to estimate the item properties of the inventories based on the GRM. These parameters may subsequently be fixed, and the model may be used to estimate a person’s level on a latent trait in a much smaller sample.
6.3 Conclusions

Given perfectionism’s role in the cause, maintenance, and course of several psychopathological states (Shafran et al., 2002; Shafran & Mansell, 2001), researchers should strive to better understand the cognitive mechanisms underlying this construct. The EEG studies by Pieters et al. (2007), Schrijvers et al. (2010), and (Tops et al., 2013), in examining the relation between the FMPS and the ERN component, the Pe component, and post-error slowing have made important contributions in this regard. However, the studies found no association between post-error slowing and inventory scores, a surprising result when operating under the assumption that post-error slowing represents increased response caution following an error, as predicted by the conflict monitoring theory (Botvinick et al., 2001).

My project expands upon the current literature by explicitly testing the key assumption of the past studies that post-error slowing is indeed representative of a more cautious mode of responding. I developed a novel quantitative model for the Simon task to test this assumption. Through the use of this model, I was able to provide more stringent controls to rule out the effects of motivation, fatigue, practice, and the possibility of multiple types of post-error adjustments. Note that aggregate measures of post-error slowing (and post-error accuracy, for that matter) would not have adequately controlled for potential confounds nor would they have the clear links and precise tests of the underlying cognitive mechanisms that my model provided.

Critically, I found that while the inclusion of a explicit mechanism for increased response caution was able to account for the slowed response times in my data, the mechanism also produced substantial overestimates of the post-error accuracy of subjects. My results are not consistent with the conflict monitoring theory, the
idea that post-error slowing is a behavioral marker of increased caution and more
cognitive control. Hence, the lack of a correlation between measures of perfectionism
and post-error slowing is not surprising, as post-error slowing is an inappropriate
marker of the mechanisms of interest. Furthermore, this suggests that interference
tasks such as the Simon task or Ericksen flanker task may not be as appropriate to
examine the cognitive mechanisms of perfectionism as originally thought.

The most important finding in this work is the fact that it is absolutely vital
to check the basic assumptions regarding what post-error slowing represents before
interpreting results in a specific theoretical framework. This sentiment has been
expressed by several researchers challenging the conflict theory account of post-error
slowing (e.g. Notebaert et al., 2009), but unfortunately applied researchers continue to
frame their designs and results based the assumption that post-error slowing is solely
reflective of a more cautious mode of responding (e.g. Spinelli et al., 2011; Luijten, van
Meel, & Franken, 2011; Shiels, Tamm, & Epstein, 2012; Sokhadze et al., 2012; Strozyk
& Jentzsch, 2012). I have shown that by relying on the unique mappings of theories
of post-error slowing onto specific sequential sampling parameters, as proposed by
Dutilh, Vandekerckhove, et al. (2012), one can establish with much more precision
and reassurance an appropriate theoretical framework to interpret results, regardless
of whether the study is using purely behavioral measures or taking advantage of EEG
methods as well. Future research into the cognitive mechanisms of perfectionism
using simple choice and response time tasks, or any clinical study making use of such
tasks, could greatly benefit from this methodology.
Appendix A: Model Fits for all Subjects

This appendix contains plots of the posterior predictive checks for all 29 subjects. The posterior predictive checks include

• the quantile-probability plots of correct and error responses for each congruency condition, where the ellipses represent the 95% credible intervals for the model predictions (pages 183 to 192);

• the posterior predictive checks of subjects’ accuracy per block, with the model’s predictions regarding the median accuracy and its associated 95% credible intervals denoted by gray lines (pages 193 to 200);

• the posterior predictive checks of a subject’s overall response times per block, with the grey bands denoting 95% credible intervals for the model predictions (pages 201 to 209); and

• the quantile-probability plots of correct and error response time distributions for trials following an error, where the ellipses represent the 95% credible intervals for the model predictions (pages 210 to 217).
Figure A.1: Quantile-probability plots of correct and error responses per congruency condition for Subject 1, 9, and 11. Ellipses are the 95% credible intervals for model predictions. The model displayed good fit to the data.
Figure A.2: Quantile-probability plots of correct and error responses per congruency condition for Subject 14, 19, and 24. Ellipses are the 95% credible intervals for model predictions. The model displayed good fit to the data.
Figure A.3: Quantile-probability plots of correct and error responses per congruency condition for Subject 26, 27, and 29. Ellipses are the 95% credible intervals for model predictions. The model displayed good fit to the data.
Figure A.4: Quantile-probability plots of correct and error responses per congruency condition for Subject 2, 3, 8, and 10. Ellipses are the 95% credible intervals for model predictions. The model displayed adequate fit to the data.
Figure A.5: Quantile-probability plots of correct and error responses per congruency condition for Subject 14, 19, and 24. Ellipses are the 95% credible intervals for model predictions. The model displayed adequate fit to the data.
Figure A.6: Quantile-probability plots of correct and error responses per congruency condition for Subject 4, 5, 6, and 7. Ellipses are the 95% credible intervals for model predictions. The model displayed poor fit to the data.
Figure A.7: Quantile-probability plots of correct and error responses per congruency condition for Subject 13, 15, and 16. Ellipses are the 95% credible intervals for model predictions. The model displayed poor fit to the data.
Figure A.8: Quantile-probability plots of correct and error responses per congruency condition for Subject 17, 18, and 20. Ellipses are the 95% credible intervals for model predictions. The model displayed poor fit to the data.
Figure A.9: Quantile-probability plots of correct and error responses per congruency condition for Subject 17, 18, and 20. Ellipses are the 95% credible intervals for model predictions. The model displayed poor fit to the data.
Figure A.10: Quantile-probability plots of correct and error responses per congruency condition for Subject 21, 23, and 29. Ellipses are the 95% credible intervals for model predictions. The model displayed poor fit to the data.
Figure A.11: Accuracy per block and session for Subject 1, 2, 3, and 4. Gray lines represent the median and lower and upper limits of the 95% credible intervals for the model predictions.
Figure A.12: Accuracy per block and session for Subject 5, 6, 7, and 8. Gray lines represent the median and lower and upper limits of the 95% credible intervals for the model predictions.
Figure A.13: Accuracy per block and session for Subject 9, 10, 11, and 12. Gray lines represent the median and lower and upper limits of the 95% credible intervals for the model predictions.
Figure A.14: Accuracy per block and session for Subject 13, 14, 15, and 16. Gray lines represent the median and lower and upper limits of the 95% credible intervals for the model predictions.
Figure A.15: Accuracy per block and session for Subject 17, 18, 19, and 20. Gray lines represent the median and lower and upper limits of the 95% credible intervals for the model predictions.
Figure A.16: Accuracy per block and session for Subject 21, 22, and 23. Gray lines represent the median and lower and upper limits of the 95% credible intervals for the model predictions.
Figure A.17: Accuracy per block and session for Subject 24, 25, and 26. Gray lines represent the median and lower and upper limits of the 95% credible intervals for the model predictions.
Figure A.18: Accuracy per block and session for Subject 27, 28, and 29. Gray lines represent the median and lower and upper limits of the 95% credible intervals for the model predictions.
Figure A.19: The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles by block for Subject 1, 2, 3, and 4. Gray bands represent the 95% credible intervals for model predictions regarding the median (dark gray), 0.3 and 0.7 (medium gray) and 0.1 and 0.9 (light gray) quantiles.
Figure A.20: The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles by block for Subject 5, 6, 7, and 8. Gray bands represent the 95% credible intervals for model predictions regarding the median (dark gray), 0.3 and 0.7 (medium gray) and 0.1 and 0.9 (light gray) quantiles.
Figure A.21: The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles by block for Subject 9, 10, 11, and 12. Gray bands represent the 95% credible intervals for model predictions regarding the median (dark gray), 0.3 and 0.7 (medium gray) and 0.1 and 0.9 (light gray) quantiles.
Figure A.22: The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles by block for Subject 13, 14, 15, and 16. Gray bands represent the 95% credible intervals for model predictions regarding the median (dark gray), 0.3 and 0.7 (medium gray) and 0.1 and 0.9 (light gray) quantiles.
Figure A.23: The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles by block for Subject 13, 14, 15, and 16. Gray bands represent the 95% credible intervals for model predictions regarding the median (dark gray), 0.3 and 0.7 (medium gray) and 0.1 and 0.9 (light gray) quantiles.
Figure A.24: The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles by block for Subject 17, 18, 19, and 20. Gray bands represent the 95% credible intervals for model predictions regarding the median (dark gray), 0.3 and 0.7 (medium gray) and 0.1 and 0.9 (light gray) quantiles.
Figure A.25: The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles by block for Subject 21, 22, and 23. Gray bands represent the 95% credible intervals for model predictions regarding the median (dark gray), 0.3 and 0.7 (medium gray) and 0.1 and 0.9 (light gray) quantiles.
Figure A.26: The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles by block for Subject 24, 25, and 26. Gray bands represent the 95% credible intervals for model predictions regarding the median (dark gray), 0.3 and 0.7 (medium gray) and 0.1 and 0.9 (light gray) quantiles.
Figure A.27: The 0.1, 0.3, 0.5, 0.7, and 0.9 response time quantiles by block for Subject 27, 28, and 29. Gray bands represent the 95% credible intervals for model predictions regarding the median (dark gray), 0.3 and 0.7 (medium gray) and 0.1 and 0.9 (light gray) quantiles.
Figure A.28: Quantile-probability plots of correct and error response times for trials following an error for Subject 1, 2, 3, and 4. Ellipses are the 95% credible intervals for model predictions.
Figure A.29: Quantile-probability plots of correct and error response times for trials following an error for Subject 5, 6, 7, and 8. Ellipses are the 95% credible intervals for model predictions.
Figure A.30: Quantile-probability plots of correct and error response times for trials following an error for Subject 9, 10, 11, and 12. Ellipses are the 95% credible intervals for model predictions.
Figure A.31: Quantile-probability plots of correct and error response times for trials following an error for Subject 13, 14, 15, and 16. Ellipses are the 95% credible intervals for model predictions.
Figure A.32: Quantile-probability plots of correct and error response times for trials following an error for Subject 17, 18, 19, and 20. Ellipses are the 95% credible intervals for model predictions.
Figure A.33: Quantile-probability plots of correct and error response times for trials following an error for Subject 21, 22, and 23. Ellipses are the 95% credible intervals for model predictions.
Figure A.34: Quantile-probability plots of correct and error response times for trials following an error for Subject 24, 25, and 26. Ellipses are the 95% credible intervals for model predictions.
Figure A.35: Quantile-probability plots of correct and error response times for trials following an error for Subject 27, 28, and 29. Ellipses are the 95% credible intervals for model predictions.
References


218


223


Psychosomatic medicine, 69(3), 249–255.