MODELS AND DYNAMICAL ANALYSES OF NEURAL SYSTEMS FOR THE ERIKSEN DECISION TASK

YUAN LIU

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
PHYSICS

SEPTEMBER, 2007
Abstract

In this dissertation I focus on the Eriksen task, a two-alternative forced choice task in which subjects must correctly identify a central stimulus and discard flankers that may or may not be compatible with it. I carry out dynamical system analysis on both a neural network model and a Bayesian inference model for it. I also extend the Bayesian inference model to study and simulate sequential effects in the Eriksen task. The scientific motivation is to understand neural mechanisms that underlie decision making tasks, including how conflicting flanker stimuli can interfere with information processing of the central stimulus, and how subjects allocate attention during task performance.

First, I study and analyze a connectionist model for the Eriksen flanker task. When solutions remain within the central domain of the logistic activation function, I show that analytical solutions of a decoupled, linearized model that is modulated by a pre-determined attention signal can provide reasonable estimates of behavioral data. I also show that the dynamics of the two-unit decision layer can be decoupled and reduced to a drift-diffusion model (DDM) with a variable drift rate, that describes the accumulation of net evidence in favor of one or the other alternative. I compare my results with numerical simulations of the full nonlinear model and with empirical data, and show that my results have a better fit, and use fewer parameters than the original model.

Two Bayesian inference models have been developed recently to model the Eriksen task. Both models are nonlinear, coupled discrete-time dynamical systems. I analyze the dynamics of those models by considering simplified, approximate systems that are linear and decoupled. I also investigate the continuum limits of these simplified dynamical systems,
demonstrating that Bayesian updating is closely related to a DDM.

In order to study and analyze sequential trial-to-trial effects during performance of the Eriksen task, I propose a simple extension to the Bayesian inference model. Simulations of my extension model agree with the findings of human subject experiment. Lastly, I also show that the experimental data provide evidence of prior updating during trials.
Acknowledgements

First, I would like to express my deepest gratitude to my thesis advisor Prof. Philip Holmes. He introduced me into the world of non-linear dynamical systems and computational neuroscience, and brought me the opportunity to conduct exciting scientific research in the frontiers of neuroscience. I have benefited not only from his knowledge and great insight as an advisor and a scientist, but also from his kindness and patience as a human being. With his guidance, I also made huge progress in scientific writing. What Phil taught me in the completion of this thesis and a few other journal articles will definitely continue to benefit me in my future career.

I am also grateful to Prof. Bill Bialek as my Physics Department mentor and my second advisor at Princeton. He introduced me to biophysics, and from the very first stage helped me engage with my thesis projects. He also continued on giving me feedback as I grew more experienced. Furthermore, he provided an outstanding example for how to live happily and to enjoy the company of the people around you.

I would also like to thank Prof. Jonathan Cohen. His inspiration and energy impressed me very much. It has been quite an experience to discuss questions in neuroscience and psychology with him, and his insights of how to target, motivate, and study problems have led me to think about some of the most interesting questions in neuroscience.

Part of my thesis work was done on Bayesian inference models. This would not have been possible without the help of Dr. Angela Yu. Working with Angela was such a good experience, and I learned a lot from her wide knowledge, profound intuition, and fast reactions.
I would also like to thank people that I had the pleasure of interacting with in the past few years. I thank former and current post-doctoral members in our group: Dr. Sam McClure, Dr. Tyler McMillen, Dr. Yael Niv, Dr. Patrick Simon, and Dr. Kong-Fatt Wong for their great help and suggestions in my research. I also thank the fellow students in our group: Dr. Eric Brown, Philip Eckhoff, Juan Gao, and Mike Todd, for their mutual support in pursuing our degrees.

Finally, I thank my family for their constant love and support, which was the ultimate drive for me to complete this thesis. My parents Honglu Liu and Weiguang Fan, have always been proud of me, and I am proud of them. To them I dedicate this thesis.
# Contents

Abstract iii

Acknowledgements v

Contents vii

List of Figures x

1 Introduction 1

2 Dynamical Systems Analysis of a Neural Network Model for the Eriksen Task 8
   2.1 Connectionist models and activation function ................................. 8
   2.2 A connectionist model for the Eriksen task ................................... 10
   2.3 Decoupling, linearization, and the drift-diffusion process ..................... 12
   2.4 Simulations with sigmoidal activation functions ................................ 16

3 A Drift Diffusion Model with Variable Drift Rate 21
   3.1 DDM reductions ................................................................. 21
      3.1.1 The interrogation protocol .............................................. 24
      3.1.2 The interrogation protocol with bounded domain ...................... 26
      3.1.3 The free-response protocol ............................................. 28
   3.2 Comparisons with empirical data ................................................. 31
3.3 Discussion ................................................. 33

4 Dynamical Systems Analysis of Inference Models for the Eriksen Task 36
4.1 Bayesian inference models for neural systems ..................... 36
4.2 A Bayesian framework for the Eriksen task ......................... 37
4.3 Linearization and parametric dependence ............................ 39
  4.3.1 The compatibility bias model ................................ 40
  4.3.2 The spatial uncertainty model ................................ 44
4.4 A continuum limit ............................................. 46
  4.4.1 Approximating the denominators ................................ 46
  4.4.2 Taking the continuum limit .................................. 50
  4.4.3 Analytical approximations for the mean posteriors ............ 51
  4.4.4 Making use of explicit mean posteriors ....................... 56
4.5 Direct simulations and comparisons ................................ 58
  4.5.1 Fitting the models to data .................................... 59
4.6 Discussion ..................................................... 60

5 A Model for Sequential Effects in the Eriksen task 69
5.1 Experimental data of sequential effects in Eriksen task ............ 69
5.2 A model for sequential effects of Eriksen task ...................... 72
5.3 Comparison with empirical data .................................... 74
  5.3.1 Sequential effects of immediately previous trial ............ 74
  5.3.2 Sequential effects in different response latency bins .......... 76
  5.3.3 Sequential effects of longer response histories ............... 76
  5.3.4 Experimental evidence for prior updating .................... 78
5.4 Discussion ..................................................... 79

6 Conclusions .................................................... 83
6.1 Summary of the dissertation ....................................... 83
6.2 Open problems and future directions ................................ 85
List of Figures

1.1 Accuracy vs. RT in the Eriksen task. Human subjects respond slower and less accurately in the incompatible condition. In particular, accuracy is below chance (.50) for short RT’s, but approaches 1 for longer RT’s. (A) Reaction times are gauged by electromyographic activities (EMG); adapted from Gratton et al, 1988). (B) Adapted from (Servan-Shreiber, Bruno, Carter, & Cohen, 1998). The details of the data sets differ, but the compatibility effect, as well as the “dip” in accuracy under the incompatible condition for short-reaction trials, is obvious in both. ........................................... 3

2.1 Logistic activation functions, showing the effects of gain $g$ and bias $\beta$. Bias sets the center of the input range over which the response is approximately linear, and gain sets the size of this range. Solid blue curve: $g = 1$, $\beta = 0$; dashed red curve: $g = 2$, $\beta = 0.2$. ........................................... 9

2.2 The architecture of the Eriksen Model, showing inputs corresponding to stimulus $<$ in the center with compatible (solid) and incompatible (dashed) flanking stimuli. ........................................... 12

2.3 Analytical solutions of the linearized, noise-free Eriksen model. Decision layer input (a) and output (b) for compatible stimuli, and the same functions for incompatible stimuli (c,d). Solid blue curves indicate $i_1$ and $z_1$: correct response; dashed red curves indicate $i_2$ and $z_2$: incorrect response. Here $a = b = a_c = 1$ and crossover times $t_{ci} = 2$, $t_{co} = 3$. ........................................... 16
2.4 Simulations of the noise-free Eriksen model with logistic activation functions.

Decision layer inputs $i_j$ (a) and outputs $z_j$ (b) for compatible stimuli, and for incompatible stimuli (c,d): crossovers in (c,d) cause the dip in accuracy. Solid blue curves indicate $i_1$ and $z_1$: correct response; dashed red curves indicate $i_2$ and $z_2$: incorrect response. Stimuli are applied at $t = 0$, after units have settled at resting values. Parameters are specified in text.

2.5 Simulations of the noise-free Eriksen model with logistic activation functions showing decision layer inputs $i_1$ and outputs $z_1$ (solid and dashed blue) and $i_2$ and $z_2$ (dash-dotted and dotted red); inputs in panels (a,c) and outputs in (b,d). Panels (a,b) show effect of attention $a_c = 0.5$ (solid and dotted) and $a_c = 1$ (dash-dotted and dotted), and panels (c,d) show effect of the ratio $b/a = 1$ (solid and dotted) and $b/a = 2$ (dashed and dash-dotted). See text for discussion. Other parameters as for Figure 2.4.

3.1 Inputs to the decision units computed from a simulation of the noise-free Eriksen model logistic activation functions (black +s). Left panel: incompatible stimuli, with linear (solid blue line), quadratic (red dots) and exponential (green dashes) fits to simulation data. Right panel: compatible stimuli with an exponential fit (solid blue curve). See text for discussion.

3.2 Analytical predictions of accuracy for the drift-diffusion model under the interrogation protocol with linear (solid, blue), quadratic (dotted, red) and exponential (dashed, green) drift rates, fitted to the nonlinear simulation data of Figure 3.1 as described in the text. Parameter values for the drift-diffusion process are $\lambda = 0$ (balanced), $c = 0.3$.

3.3 Accuracy vs. RT for the bounded diffusion process with an exponential drift rate and bounds $L$ of 0.1 (red), 0.3 (blue) and 0.5 (green) bottom to top on right of figure, with noise strength $c = 0.3$. Dashed black curve shows accuracy for unbounded process. Arrows at right show theoretical accuracy limits for large RT from (3.20).
3.4 Left: Accuracy vs. RT for the full Eriksen model under the interrogation protocol (solid line) compared with analytical results for the unbounded DDM (stars). Right: Accuracy vs. RT for the full Eriksen model (solid line) under the free-response protocol compared with numerical results for the unbounded DDM with exponential inputs (stars): thresholds are $\theta = 1.0$ for full model and $\theta = \pm 0.3$ for DDM. Upper curves denote compatible trials and lower curves incompatible trials. Remaining parameters are as specified in text.

3.5 Accuracy vs. RT for the full Eriksen model (blue lines with star dots) under the free-response protocol, with RT histograms below (left, dark blue bars of each pair). Unbounded DDM accuracy results are shown for comparison in solid red with open triangles and RT histograms in right, red bars of each pair. Thresholds are $\theta = 1.1$ for full model and $\theta = \pm 0.3$ for DDM. Left panel shows compatible trials, right panel shows incompatible trials. Remaining parameters are as specified in text.

3.6 Accuracy vs. RT for the full model (top, solid) and the DDM (bottom, solid), compared with data from [43] (dashed). Standard errors of latter indicated by vertical bars. Curves for compatible stimuli lie above those for incompatible stimuli.

3.7 RT histograms from the full model (top row) and the DDM (bottom row), compared with data from [43]. Model results are shown in right, yellow bars and empirical data in left, blue bars for each 50 ms bin. Compatible trials appear in left column and incompatible trials in right column.
4.1 Simulated and analytical approximations of parameter values that produce
dips in accuracy vs. reaction time for incompatible trials. Graphs show
accuracy averaged over trials with simulated reaction times under 20 time
steps, as a function of $\beta$ for the compatibility bias model (left), and the ratio
of means $a_1/a_2$ for the spatial uncertainly model (right). Crossings with
the 0.5 accuracy line indicate numerically obtained estimates of the “true”
parameter constraints; dashed vertical lines show the approximate constraints
of Eqs. (4.17) and (4.23).

4.2 Mean values of the denominator $\langle D_t \rangle$ for compatible (blue solid) and incom-
patible (red dashed) stimuli, each averaged over $10^5$ trials. In both cases the
$\langle D_t \rangle$ rises monotonically toward its upper bound $1 + \frac{3}{\sigma^2} = 1.0370\ldots$ 

4.3 Typical distributions from which the multiplicative factors $a_{i,j} + b_{i,j} \eta(t)$ in
Eqn. (4.29) are drawn on each time step. Parameter values are $\sigma = 1.8$ (top)
and $a_1 = 0.7$, $a_2 = 0.3$, $\sigma_1 = 1.4$, $\sigma_2 = 1$ (bottom). For illustrative purposes,
standard deviations $\sigma$, $\sigma_1$, $\sigma_2$ are 20% of those used in the text to reduce
overlap of distributions.

4.4 Probability density functions in logarithmic $y$-space and the original $z$-space.
Note that the probability distribution in $y$-space is a Gaussian distribu-
tion, while transferred into $z$-space, the distribution is skewed into one with
squeezed head and extended tail.
4.5 Predictions of the full and simplified compatibility bias models in the case
that the central symbol is $S (s_2 = 1)$ and with prior compatibility bias $P(M) = 0.9$. Top left: marginal mean posterior probabilities $P(s_2 = 1|M)$ (correct response) for compatible and incompatible conditions. Top right: marginal mean posterior $P(M = 1)$ for compatibility. Bottom row: individual mean posteriors for compatible (left) and incompatible (right) trials. Results from full inference model, averaged as in Figure 4.5, shown solid and predictions of the continuum approximation (4.42-4.44) shown dashed. Keys identify individual curves.

4.6 Predictions of the full and simplified spatial uncertainty models. Top left: marginal mean posterior probabilities $P(s_2 = 1|M)$ (correct response) for compatible and incompatible conditions. Top right: marginal mean posterior $P(M = 1)$ for compatibility. Bottom row: individual mean posteriors for compatible (left) and incompatible (right) trials. Results from full inference model, averaged as in Figure 4.5, shown solid and predictions of the continuum approximation (4.42) and (4.45-4.47) shown dashed. Keys identify individual curves.

4.7 Predictions of accuracy (left) and reaction time histograms (right) computed under the approximation of Section 4.4.4. Solid curve and dark boxes indicate compatible trials; dashed curve and light boxes indicate incompatible trials.

4.8 Posterior probability $P(s_2 = 1|X_t)$ for one sample path of the approximate compatibility bias model (Eq. (4.16), dashed), compared with a sample path from the original inference model (Eq. (4.9), solid). The same sequence $x(t)$ of observations was used in both cases.
4.9 Accuracy and reaction time distributions for the compatibility bias model for compatible stimuli (left) and incompatible stimuli (right). Solid and right hand (blue) bar of each RT bin pair from full inference model of [117]; dashed and left hand (yellow) bars from approximate (linearized likelihood) model. Results obtained by averaging over 2,000 simulated trials. 66

4.10 Accuracy and reaction time distributions for the spatial uncertainty model for compatible stimuli (left) and incompatible stimuli (right). Format as in Figure 4.9. Results obtained by averaging over 2,000 simulated trials. 66

4.11 Accuracy (upper curves) and reaction time distributions (lower curves) from the full (squares) and reduced DD (triangles) models for compatible (left) and incompatible (right) trials. Upper panels show compatibility bias and lower panels spatial uncertainty model results respectively. Parameters were fitted to the data of [43] (dashed curves). 67

4.12 Accuracy (upper curves) and reaction time distributions (lower curves) from the full (squares) and reduced DD (triangles) models for compatible (left) and incompatible (right) trials. Upper panels show compatibility bias and lower panels spatial uncertainty model results respectively. Parameters were fitted to the data of [92] (dashed curves). 68

5.1 Average reaction times (left panel) and error rates (right panel) for each of four experimental conditions. Responses to compatible trials were faster if the previous trial was also compatible (by 10 ms) and responses to incompatible trials were faster (although by only 3 ms) and more accurate (error rate was 0.075 lower) if the previous trial was also incompatible. Left: mean reaction times over all trials. Right: error rates. From [44] Figure 1. 70

5.2 Conditional accuracy functions for the four experimental conditions. (Response latency is shown in milliseconds.) Experiment reveals that the sequential effect was larger after a compatible trial than after an incompatible trial. From [44] Figure 2. 71
5.3 Given equal probability of being compatible and incompatible, the prior gradually decays to 0.5. The simulation is averaged over 1000 trials, with initial biased prior $\beta = 0.9$, and $\delta = 0.05$. ....................... 73

5.4 Comparison with empirical data. Average reaction times (left panel) and error rates (right panel) for each of the experimental conditions. Top panels are experimental data, and bottom panels are simulations. ....................... 75

5.5 Simulation of conditional accuracy functions for each of the experimental conditions. (response latency is shown in 20 step bins. Compare with Figure 5.2). ....................... 77

5.6 Reaction time and accuracy as a function of preceding stimulus compatibility sequences. Top panels are experimental data, while bottom ones are simulation results. “C-I” sequences read downward, so that the current trial appears at the bottom in all cases. ....................... 78

5.7 Reaction time and accuracy as a function of preceding stimulus compatibility sequences. Blue curves: averaged value over trials in first half of all blocks. Red curves: averaged value over trials in second half of all blocks. Top panels are experimental data, while bottom ones are simulation results. ....................... 80

B.1 The integral of the joint posterior probability distribution is taken over the positive $(z_1, z_2)$-quadrant less the shaded triangular region. ....................... 90
Chapter 1

Introduction

Our daily lives are full of decision making and the study of decision making spans varied fields as neuroscience, psychology, physics, mathematics, economics, political science, and computer science. There is currently considerable interest in the psychological [63, 81, 83] and neural bases of decision making [88, 73, 39, 40]. In this dissertation, I primarily consider simple two-alternative forced-choice (2AFC) tasks and the associated neural activities in the brain. I hope that the studies of such simple tasks can be extended to more complicated ones, and that they will help to identify principles that contribute to decision making processes in higher cognition level.

In a simple 2AFC task, a subject must decide, on each trial, which of two randomly-presented stimuli is presented, and respond accordingly. The discrimination process is typically modeled as a competition among different populations of neurons, each preferentially responsive to one of the stimuli [106]. This is supported by direct recordings in oculomotor areas of monkeys performing such tasks, which suggest that “decision” neurons (e.g. in the lateral interparietal area (LIP) and frontal eye field (FEF)) accumulate evidence for the stimulus alternatives, and that the corresponding behavioral response is initiated when their firing rates cross thresholds, e.g. [88, 39, 96, 90, 86]. Moreover, computational simulations and analyses of neural network (connectionist or parallel distributed processing (PDP)) models [45, 53, 54] show that their solutions can be matched to behavioral
data [21, 23, 106].

In particular, a simple and analytically-tractable drift-diffusion model (DDM) has been extensively fitted in this manner [81, 83]. It is also known as the continuum limit of the sequential probability ratio test (SPRT), which was developed by Abraham Wald as a hypothesis test for sequential analysis [108]. The DDM is widely used in physics and mathematics. Psychologists have replaced the density distribution of gas in physics with the probability density distribution of suitable variables in psychology, such as the firing rate of a neuron.

The diffusion model was proposed as a model for decision making and memory recall by Ratcliff in 1978 [81]. He assumes that during the decision process, the difference between the amounts of evidence supporting the two choices is accumulated. The DDM is known to be optimal for 2AFC in the sense that, on average, it delivers choices of guaranteed accuracy in the shortest possible time [63], and analytical solutions for DDM error rates and decision times from the DDM can be used to investigate speed-accuracy tradeoffs for optimal performance [40, 11]. Moreover, it has been shown that various neural network models of decision-making [21, 23, 106] can be reduced to variants of the DDM [11]. However, little work has examined the ability of this model to account for performance in more complex and cognitively-interesting tasks, such as those requiring selective attention to a subset of task-relevant stimuli in the presence of distracters. The Eriksen flanker task has been used extensively to study such effects [28, 43, 23].

In this task subjects are asked to respond appropriately to a target letter or arrow (e.g. < or >), visually displayed in the center of a five-symbol stimulus array on a display screen (e.g., by pressing the left button to < and the right button to >). The flanking symbols may be either compatible or incompatible with the central stimulus. In the compatible conditions, the display reads "<<<<<< or >>>>>>; in the incompatible conditions, it reads ">>>>> or "<<<<<<, and in each block of trials all four conditions are typically presented with equal probabilities.

Experiments show that subjects are slower and make more errors under the incompati-
Figure 1.1: Accuracy vs. RT in the Eriksen task. Human subjects respond slower and less accurately in the incompatible condition. In particular, accuracy is below chance (.50) for short RT’s, but approaches 1 for longer RT’s. (A) Reaction times are gauged by electromyographic activities (EMG); adapted from (Gratton et al, 1988). (B) Adapted from (Servan-Shreiber, Bruno, Carter, & Cohen, 1998). The details of the data sets differ, but the compatibility effect, as well as the “dip” in accuracy under the incompatible condition for short-reaction trials, is obvious in both.
ble conditions, as illustrated in the data of [43]: see Figure 1.1. Furthermore, response patterns exhibit an interesting temporal profile: specifically, a dip in accuracy for incompatible trials at short reaction times, and a “crossover time” at which accuracy regains 50%: the chance level for “blind” responses. This dip is thought to reflect the dynamics of an interaction between bottom-up processing of sensory information (which, for the incompatible condition, favors the incorrect response) and the engagement of “top-down” attentional processes which favor processing of the central stimulus and thereby encourage the correct response. Accuracy for compatible trials increases monotonically with time.

Trials may be run under a free-response paradigm in which decisions are signaled when the subject feels that sufficient evidence in favor of one alternative has accumulated. Since sensory and decision processes are subject to variability, response times vary from trial to trial and performance under the free-response condition is characterized by both reaction time distributions and error rates. In contrast, in a forced-response or deadline paradigm, subjects must respond at or before a fixed time $T$ following stimulus onset with their best estimate of which alternative was presented. This is how Figure 1.1 was generated. Reaction times may still vary due to errors in temporal estimation. Here I shall consider both free response and the “hard limit” of forced response, in which the decision must be rendered when a cue is given: in this limiting case I can ignore RT variability and consider only accuracy as a function of the cue time. To distinguish the latter from deadlining, I call it the interrogation protocol. In both cases one can sort the data into response time bins and plot it as in Figure 1.1, but as it shall be seen, the two cases lead to somewhat different predictions.

Cohen et al. [23] proposed a neural network model of the Eriksen task and showed that it can be fitted to the data of [43]. The model has multiple layers and includes top-down biases applied to perception units associated with the central stimuli. However, like other connectionist models with nonlinear input-output response functions, it is not amenable to analysis, and data fitting and predictive studies must be carried out by numerical simulation, cf. [92, 93]. In this dissertation I derive a simplified, linearized system and show how it can
be reduced to a DDM with variable drift rate that models the decision process. In doing so I derive analytical approximations for crossover times and other characteristics that assist parameter fitting, and reveal how the DDM emerges naturally from more complex multi-layered networks.

On the other hand, since the Eriksen task is an extension of the standard 2AFC task, there are reasons to suspect that the optimal policy in this case should be similar to the SPRT. In this vein, one recent work [117] modeled the computations underlying the Eriksen task as iterative Bayesian updating, and the perceptual decision as being made (and the trial terminated) when the cumulative posterior for one of the two possible settings of the target stimulus exceeds a fixed decision threshold. It was also proposed that the apparent suboptimality in the subjects’ performance can be explained by either an incorrect prior on the relative frequency of compatible and incompatible trials (compatibility bias model), or by inherent spatial corruption in the generation of visual inputs (spatial uncertainty model) [117]. Building on that work, I approximate the Bayesian inference models with simpler dynamical systems, and study them both analytically and via simulations. I demonstrate that behavioral performance in the approximate models closely follows the original models, but that their analytical tractability allows one to derive explicit expressions for the dependence of inferential and psychometric quantities on model parameters. My analysis reveals the formal similarity of computations underlying the compatibility bias and spatial uncertainty models, which were motivated by disparate experimental literature and were formulated differently within the Bayesian framework. I also discuss the relationship between exact Bayesian inference and drift-diffusion processes, which are continuum limits of the approximate models.

Bayesian inference models provide a new way to study cognitive behaviors, but problems still remain. For example, the prior in the inference model is fixed, and the process is Markovian: the current state depends only on the previous state. Then how the memory stores and retrieves is crucial. One of the memory effects is the sequential effect. Sequential effects have been observed widely in cognitive experiments, especially in decision making.
tasks. A considerable number of experiments have demonstrated that reaction time (RT) to a stimulus is dependent not only on that stimulus but also on the sequence of preceding stimuli [101, 102]. These aftereffects of previous stimulus-response cycles are called sequential effects. The first-order sequential effect is the RT difference caused by the immediate preceding stimulus-response cycle, whereas the effects caused by stimuli earlier in the sequence are called higher order sequential effects. The first concept that may account for sequential effects stems from strategies and expectancies. It is known that people have an irresistible tendency to expect particular events even when the events are generated randomly. In this way, subjects expect the appearance of a particular stimulus in an RT experiment and base their preparation upon that expectation. I propose a very simple model to simulate sequential effects in the Eriksen task. I show that simulations agree with experimental findings, and experimental data also provide evidence for prior updating from trial to trial.

This dissertation is organized as follows. In Chapter 2 I briefly review connectionist network models, describe the model of [23], and analyze a linearized and decoupled version of it, which finally results in a DDM describing the evolution of net evidence in favor of one (or the other) alternative. A key ingredient is the time-varying input (drift) to the DDM from the perception and attention layers of the network, and I compare my analytical predictions of this with numerical simulations of the original nonlinear system. Chapter 3 contains an analysis of the DDM with drift rates of various functional forms derived from simulation data. I compute accuracy vs response time curves for the interrogation protocol explicitly (displaying parameter dependencies), and for the free response protocol numerically, and compare them with simulations of the full Eriksen model. I also compare fits of the original model of [23] and the reduced DDM to empirical data of [43]. In Chapter 4, I review the inference models, and derive and analyze simplified models: uncoupled, linear discrete dynamical systems. From these I derive the continuum limit, a stochastic ordinary differential equation (SDE), which, in logarithmic probability space, is also a DDM. From this, I compute analytical predictions for mean trajectories of the posterior probabilities. I
also compute accuracy vs. time curves and reaction time distributions under an approximation that violates the first passage threshold crossing criterion adopted in [117], but permits explicit analysis. The behavioral data for sequential effects of Eriksen tasks is reviewed in Chapter 5, followed by a Bayesian model for it. By changing compatibility bias from trial to trial, I simulate the changes of reaction time and accuracy given different stimuli combinations of previous trials. The model is also fit to experimental data and the data provide supporting evidence for prior adaptation in continuous trials. Chapter 6 contains summaries, conclusions, and proposals for future work.

This dissertation contributes to the mathematical and computational modeling of decision making and cognition. The major contributions are demonstrating that both biophysically realistic multi-layer neural network models and normative Bayesian models of decision making can be reduced to scalar drift-diffusion processes. This yields analytical approximations for behavioral observables in terms of model parameters, and provides mechanistic explanations of simple cognitive processes. The fitting error calculation and analysis show that the accuracy of the reduction is comparable with original models, although the reduction is computationally much simpler and has fewer fitting parameters. Furthermore, the mathematical analysis relating to updating of posterior probabilities in Bayesian decision methods to drift-diffusion models is the first of its kind in understanding of the dynamics of stimulus recognition and decision making tasks. It could have a significant impact on connecting two kinds of neural models, and understanding of brain functions and neuron activities of probability estimation in decision making tasks, or more generally, any other cognitive tasks. This dissertation also suggests stochastic differential equations (SDEs) as a modeling and analysis tool to study neural network models, and provides analytical solutions for some specific cases. This could help to simplify more complicated and higher order cognitive processes in the future.
Chapter 2

Dynamical Systems Analysis of a Neural Network Model for the Eriksen Task

2.1 Connectionist models and activation function

Connectionist models are stochastic differential equations (SDEs) or iterated mappings. The neural network branch of connectionism suggests that the study of brain activity is really the study of neural systems. This area links connectionist models to neuroscience, and offers models having varying degrees of biological realism, for example, units in the network could represent neurons and the connections could represent synapses. In connectionist models, the activities (firing rates) of neurons or groups of neurons evolve in a manner determined by their summed inputs transformed via an activation or response function:

$$\text{output} = \psi \left( \sum_{i=1}^{n} w_i x_i + I \right),$$  \hspace{1cm} (2.1)

where $\psi$ is bounded below and above to reflect the resting and maximal firing rates. Here $x_i$ are the activities of other neurons or groups connected via weights $w_i$, negative values representing inhibition and positive values excitation. The term $I$ models external inputs.
from stimuli, perhaps modulated via sensory circuits. Activation functions for the neural
network are needed to introduce nonlinearity into the network. The reason is that a linear
function of linear functions is again a linear function. However, it is the nonlinearity (i.e,
the capability to represent nonlinear functions) that makes multilayer networks so powerful.
Typically $\psi$ is taken as the sigmoid

$$\psi(x) = \frac{1}{1 + \exp(-4g(x - \beta))},$$

with parameters $g$ and $\beta$ specifying gain and bias. Bias sets the input range $x \approx \beta$ in which
the unit is maximally responsive, and gain determines its width. Outside this region, the
unit is essentially quiescent (output = 0) or maximally active (output = 1): see Figure 2.1.
Appropriate parameter choices can effectively increase signal-to-noise ratios as information
flows though a network, by amplifying larger inputs and suppressing smaller ones [94]. Eqn.
(2.2) is chosen so that the maximal slope $\partial \psi/\partial x(\beta) = g$, and it has been argued that neural
circuits should adjust biases $\beta$ to work near this point to utilize the resulting sensitivity
[21]. As in earlier work [17], we shall appeal to this in linearizing response functions at their
maximal slope regions to yield more tractable models.

![Logistic activation functions](image)

**Figure 2.1:** Logistic activation functions, showing the effects of gain $g$ and bias $\beta$. Bias sets
the center of the input range over which the response is approximately linear, and gain sets
the size of this range. Solid blue curve: $g = 1$, $\beta = 0$; dashed red curve: $g = 2$, $\beta = 0.2$.

Models such as those considered below can reproduce and help explain a wide range of
behavioral data [22, 92, 93, 7], and moreover may be derived from biophysically-based ionic current models of single cells [1]. Since the latter take the form of continuous differential equations rather than iterated maps, SDEs shall be focused in this paper.

2.2 A connectionist model for the Eriksen task

I consider the architecture proposed by [23], shown in Figure 2.2. The units \( p_1 \) through \( p_6 \) constitute the perception layer (called the input module in [23]), \( z_1 \) and \( z_2 \) constitute the decision layer (output module), and \( a_1 \) through \( a_3 \) the attention module. Units within each layer or module are mutually interconnected via inhibitory weights \( -w \) (here all assumed equal), implementing competition among representations within that layer. The decision and perception layers receive excitatory inputs of weights \( +l \) and \( +h \) from the perception and attention layers respectively, and the left, center and right units of the attention layer receive excitatory inputs of weights \( +h \) from the corresponding pairs of perception units, as shown. All units are subject to independent additive white noise, to simulate unmodeled inputs. Absent inputs and noise, each unit’s activity decays at rate \( -k \).

The perception layer contains three pairs of units that receive inputs from the left, central, and right visual fields respectively and it is assumed that in each pair the left unit is preferentially responsive to the symbol \(<\) and the right unit to \(>\). Thus, stimuli are modeled as follows: in the compatible condition, either the left \((p_3)\) or right \((p_4)\) central unit has external input \(I_j = a\), and the corresponding flanker units \((p_1, p_5 \text{ or } p_2, p_6 \text{ resp.})\) receive input \(I_j = b\) (modeling \(<<<<<<\text{ or }>>>>>>\)); and, in the incompatible condition, either the left \((p_3)\) or right \((p_4)\) central unit has external input \(I_j = a\), and the non-corresponding flanker units \((p_2, p_6 \text{ or } p_1, p_5 \text{ resp.})\) receive input \(I_j = b\) (modeling \(>>><<\text{ or }<<<<\)). Since each flanker unit represents two symbols in the stimulus array, typically assume \(b \geq a\).

The central unit \((a_2)\) of the attention layer receives an input \(A_2 = a_c\) in both conditions. All other inputs \(I_j, A_j\) are zero. See Figure 2.2 for an example.

Under the interrogation paradigm the decision is rendered at a set time \(t\) after stimulus onset by taking the larger of the decision unit outputs: i.e. “\(<\)” if \(z_1(t) > z_2(t)\) and “\(>\)” if
$z_2(t) > z_1(t)$. Under the free-response paradigm the first of $z_1(t)$ or $z_2(t)$ to cross a preset threshold $z_j = \theta$ determines the choice.

In this Chapter I consider continuously evolving SDE models of these mechanisms, which for the Eriksen model may be written as:

\[
\begin{align*}
\dot{z}_1 &= -kz_1 + \psi(-wz_2 + l(p_1 + p_3 + p_5)) + \eta_{a1}, \\
\dot{z}_2 &= -kz_2 + \psi(-wz_1 + l(p_2 + p_4 + p_6)) + \eta_{a2}; \\
\dot{p}_1 &= -kp_1 + \psi(-w(p_2 + p_3 + p_4 + p_5 + p_6) + ha_1 + I_1) + \eta_{d1}, \\
\dot{p}_2 &= -kp_2 + \psi(-w(p_1 + p_3 + p_4 + p_5 + p_6) + ha_1 + I_2) + \eta_{d2}, \\
\dot{p}_3 &= -kp_3 + \psi(-w(p_1 + p_2 + p_4 + p_5 + p_6) + ha_2 + I_3) + \eta_{d3}, \\
\dot{p}_4 &= -kp_4 + \psi(-w(p_1 + p_2 + p_3 + p_5 + p_6) + ha_2 + I_4) + \eta_{d4}, \\
\dot{p}_5 &= -kp_5 + \psi(-w(p_1 + p_2 + p_3 + p_4 + p_6) + ha_3 + I_5) + \eta_{d5}, \\
\dot{p}_6 &= -kp_6 + \psi(-w(p_1 + p_2 + p_3 + p_4 + p_5) + ha_3 + I_6) + \eta_{d6}; \\
\dot{a}_1 &= -ka_1 + \psi(-w(a_2 + a_3) + h(p_1 + p_2) + A_1) + \eta_{a1}, \\
\dot{a}_2 &= -ka_2 + \psi(-w(a_1 + a_3) + h(p_3 + p_4) + A_2) + \eta_{a2}, \\
\dot{a}_3 &= -ka_3 + \psi(-w(a_1 + a_2) + h(p_5 + p_6) + A_3) + \eta_{a3},
\end{align*}
\]  
(2.3-2.5)

where the $\eta_j$’s represent i.i.d. white noise processes.

This 11-dimensional, coupled set of SDEs is effectively insoluble analytically, so I shall employ two strategies that result in more tractable approximations: linearization and decoupling. In all, with the stimulus and attention input choices specified above, 11 parameters are required to specify the system ($g$ and $\beta$ for the sigmoids; $w$, $l$, $h$ and $k$ for leak and connection weights, $a$, $b$ and $a_c$ for inputs, a threshold value, and an overall noise level.) Allowing different values in each layer would significantly increase this number. The reduction carried out below substantially reduces the number of parameters. Similar analyses of simpler models of the 2AFC task are developed in [17] and [11].
2.3 Decoupling, linearization, and the drift-diffusion process

Here I use two major ideas to simplify the problem: decoupling and linearization. I decouple the lower layers of the model by assuming that the modulatory output of the attention layer to the perception layer has a predetermined time course that is little-affected by feedback from the perception layer (the decision layer is already decoupled in the version of the model given above, in that it does not feed back to the lower layers). I then appeal to the proposal of [21]: that biases in the sigmoidal units are adjusted so that they remain near their most sensitive ranges (close to maximum slope) where input modulations have maximal effect on outputs, and I replace the nonlinear functions (2.2) by their linearizations at $x = b$.

As already noted, the decision layer does not feed back to the perception or attention layers and thus cannot influence their dynamics. It may therefore be analyzed independently, given knowledge of, or assumptions regarding, the inputs $i_1 = l(p_1 + p_3 + p_5)$ and $i_2 = l(p_2 + p_4 + p_6)$ to its two units. Furthermore, assuming that the sigmoid bias parameter $\beta$ in (2.2) is selected so that the units remain near their most sensitive range, I linearize...
(2.3) about \( z_1 = z_2 = \beta \) and let \( \tilde{z}_j = z_j - \beta \) to obtain:

\[
\begin{align*}
\dot{\tilde{z}}_1 &= -k\tilde{z}_1 - gw\tilde{z}_2 + g\,i_1 + \eta_{a1}, \\
\dot{\tilde{z}}_2 &= -k\tilde{z}_2 - gw\tilde{z}_1 + g\,i_2 + \eta_{a2}.
\end{align*}
\]  

(2.6)

Subtracting these equations yields a scalar Ornstein-Uhlenbeck process:

\[
\dot{u} = (gw - k)u + A + \eta,
\]  

(2.7)

where \( u = \tilde{z}_1 - \tilde{z}_2 \) and \( A = g(i_1 - i_2) \) is the difference in the inputs. If \( A > 0 \) \((i_1 > i_2)\) \( u \) will tend to increase and if \( A < 0 \) \((i_1 < i_2)\) it will tend to decrease: thus, in its linearized form, the decision layer integrates the net evidence from the perception layer.

When \( gw - k = 0 \), the first term on the right of equation (2.7) vanishes, and I say that such a network is balanced [11]. In this case (2.7) is a pure drift-diffusion process, and is particularly simple to analyze, even when the net evidence \( A(t) \) varies with time, as it will in Chapter 3.

I will also formally decouple the perception layer from the attention layer, assuming that the feedback from the latter may be approximated by a specified time-dependent function. I initially neglect noise and inputs from the attention layer, so that after linearization, setting \( \bar{p}_j = p_j - \beta \), and writing \( p = (\bar{p}_1, \ldots, \bar{p}_6) \), I have the linear ODE system:

\[
\dot{p} = A\,p + I,
\]  

(2.8)

where the \( n \times n \) matrix \( A \) and input vector \( I \) are:

\[
A = \begin{bmatrix}
-k & -gw & -gw & -gw & -gw & -gw \\
-gw & -k & -gw & -gw & -gw & -gw \\
-gw & -gw & -k & -gw & -gw & -gw \\
-gw & -gw & -gw & -k & -gw & -gw \\
-gw & -gw & -gw & -gw & -k & -gw \\
-gw & -gw & -gw & -gw & -gw & -k
\end{bmatrix}, \quad I = \begin{bmatrix}
b \\
a \\
a \\
0 \\
b \\
0
\end{bmatrix}
\]  

(2.9)

Here I assume that the central stimulus is "<" and the components of the vector \( I \) respectively correspond to compatible and incompatible conditions. The ">" case may be
derived using symmetry arguments. Also note that, since the inputs to the flanker units $p_1, p_5$ and $p_2, p_6$ are equal in both compatible and incompatible cases, solutions remain on the invariant 4-dimensional plane $p_1 = p_5, p_2 = p_6$, provided that the corresponding initial conditions are also equal.

The eigenvalues of the symmetric matrix $A$ are $\lambda_1 = -(k+5gw)$ and $\lambda_2 = -(k-gw)$ with multiplicities 1 and 5 respectively, so I may diagonalize (2.8) by an orthogonal transformation $p = Ty$, solve the resulting decoupled ODEs in the $y$ coordinates and transform back to $p$, as detailed in Appendix A. In this way I may compute the sums that form the inputs to the decision layer in (2.3) and its linearization (2.7):

$$i_{1,2} = \frac{3y_1}{\sqrt{6}} \pm \left[ \frac{2y_2}{\sqrt{2}} - \frac{y_3}{\sqrt{6}} + \frac{2y_4}{\sqrt{12}} \right]; \quad (2.10)$$

in writing (2.10) I have also used the symmetry $p_1 = p_5, p_2 = p_6$.

If $y_j(0) = 0$, corresponding to unbiased starting points, and $a, b = \text{constant}$, the general solution given in Appendix A yields:

$$i_{1,2} = \frac{3y_1}{\sqrt{6}} \pm \left[ \frac{2y_2}{\sqrt{2}} - \frac{y_3}{\sqrt{6}} + \frac{2y_4}{\sqrt{12}} \right]; \quad (2.11)$$

In (2.11) the central $\pm$ refers to the cases $i_1, i_2$ and the right-hand $\pm$ to the compatible/incompatible conditions. The special case in which decay and inhibition are balanced $(k = gw, \lambda_1 = -(1 + 5g)k \overset{\text{def}}{=} -\bar{\lambda}, \lambda_2 = 0)$ is of particular importance:

$$i_{1,2} = \frac{3y_1}{\sqrt{6}} \pm \left[ \frac{2y_2}{\sqrt{2}} - \frac{y_3}{\sqrt{6}} + \frac{2y_4}{\sqrt{12}} \right]; \quad (2.12)$$

Since the difference between the inputs to the decision layer is

$$i_1 - i_2 = \frac{l(a \pm 2b)(e^{\lambda_2 t} - 1)}{\lambda_2} \quad (\text{or } l(a \pm 2b)t \text{ when } \lambda_2 = 0), \quad (2.13)$$

the flanker inputs dominate in the incompatible case, as one expects (provided $b > a/2$). However, the inputs need not remain constant: in [23] and [92] the central perception units $p_3, p_4$ are activated on all trials via the output of the set of attention units shown in Figure 2.2. For simplicity I shall initially model this effect by boosting the central inputs
$I_{3,4}$ to the perception layer by a multiplicative factor that increases linearly with time, replacing $a$ thus:

$$a \mapsto (1 + a_c t)a.$$ \hfill (2.14)

I shall partially justify and improve this simple choice subsequently by comparing with the decision layer inputs and outputs in the nonlinear model (2.3-2.5). Replacing $a$ as in (2.14) and using the solutions given in the Appendix A adds the following terms to the individual inputs:

$$i_{1,2} = \ldots + laa_c \left[ \frac{(e^{\lambda_1 t} - 1 - \lambda_1 t)}{2\lambda_1^2} \pm \frac{(e^{\lambda_2 t} - 1 - \lambda_2 t)}{2\lambda_2^2} \right],$$ \hfill (2.15)

so that the differences become

$$i_1 - i_2 = l \left[ \frac{(a \pm 2b)(e^{\lambda_2 t} - 1)}{\lambda_2} + aa_c \frac{e^{\lambda_2 t} - 1 - \lambda_2 t}{\lambda_2^2} \right],$$ \hfill (2.16)

and

$$i_1 - i_2 = l \left[ (a \pm 2b)t + \frac{aa_c t^2}{2} \right]$$ \hfill (2.17)

in the balanced case ($\lambda_2 = 0$). The inputs are equal at $t = 0$, and $i_1 > i_2$ for $t > 0$ in the compatible case, but I see that there is now a critical crossover time such that $i_1 < i_2$ for $0 < t < t_{ci}$ and $i_1 > i_2$ for $t > t_{ci}$ in the incompatible case. With balanced parameters, I have

$$t_{ci} = \frac{2(2b - a)}{aa_c},$$ \hfill (2.18)

and solving the noise-free drift equation (2.7) with $A = g(i_1 - i_2)$ of (2.17) and $u(0) = 0$, I find that

$$u(t) = l \left[ \frac{(a \pm 2b)t^2}{2} + \frac{aa_c t^3}{6} \right].$$ \hfill (2.19)

Hence for incompatible stimuli the output of the decision layer is negative for $0 < t < t_{co}$ and positive thereafter, and the crossover time for that output is given by

$$t_{co} = \frac{3(2b - a)}{aa_c}.$$ \hfill (2.20)

Figure 2.3 shows examples of decision layer inputs and outputs in the balanced case $\lambda_2 = 0$.

If the effect of attention is modeled by any additive term applied equally to both central decision units, so that in place of (2.14) I have $a \mapsto a + a_c(t)$ applied to $p_3$ and $0 \mapsto a_c(t)$...
Figure 2.3: Analytical solutions of the linearized, noise-free Eriksen model. Decision layer input (a) and output (b) for compatible stimuli, and the same functions for incompatible stimuli (c,d). Solid blue curves indicate $i_1$ and $z_1$: correct response; dashed red curves indicate $i_2$ and $z_2$: incorrect response. Here $a = b = a_c = 1$ and crossover times $t_{ci} = 2$, $t_{co} = 3$.

applied to $p_4$, I find that the difference between the inputs to the decision layer is unaffected. A nonlinear (multiplicative) interaction is evidently crucial. As one shall see, such an interaction emerges naturally from the nonlinear activation functions $\psi$ of Figure 2.1.

### 2.4 Simulations with sigmoidal activation functions

I now return to the the more neurally-realistic network (2.3-2.5) with nonlinear activation functions, and perform numerical simulations to validate the linearized analysis of §2.3. In order to make direct comparisons, I exclude noise terms again. Here and in Chapter 3 I
set parameters $k = w = l = h = 1$ and $a = b = 0.5$, $a_c = 1$ unless specified otherwise, and take sigmoidal gain and bias $g = 0.55$, $\beta = 0.8$ for the attention and perception units and $g = 1$, $\beta = -0.9$ for the decision units. The biases are selected so that, at rest without stimulus inputs, the units are close the centers of their sensitive ranges where $\psi' = g$, and I allow the units to equilibrate after starting with zero initial conditions before the stimuli are applied at $t = 0$. In next chapter I shall derive parameters by fitting to the data of Figure 1.1.

Examples of solutions analogous to those of Figure 2.3 are shown in Figure 2.4, illustrating that the linearized analysis captures the key qualitative effects exhibited by the nonlinear system following stimulus onset, including crossover behavior in incompatible cases. In particular, the inputs from the central attention unit $a_2$ via the sigmoidal function have the effect of boosting the central stimulus inputs as assumed in my simple analysis. The quantitative predictions of the linear analysis are also adequate: in particular, the ratios between decision layer input and output crossover times fall within $10 - 15\%$ of the value $r = t_{ci}/t_{co} = 2/3$ from (2.18-2.20), and the linear dependence of $t_{ci}$ and $t_{co}$ on $(2b - a)$ and inverse dependence on $a_c$ of (2.18-2.20) are approximately borne out. Figure 2.5 shows solutions for incompatible cases for different values of the attention parameter $a_c$ and of the ratio $b/a$, which measures the relative strength of flanker to center stimuli. Increasing $a_c$ by a factor of 2 reduces $t_{ci}$ from 4.43 to 1.84 and $t_{co}$ from 6.11 to 2.75, and increasing $b/a$ by 2 increases $t_{ci}$ from 1.41 to 5.33 and $t_{co}$ from 2.21 to 7.25, giving 0.42, 0.45 and 3.78, 3.28 respectively, in comparison to the factors 0.5 and 3 predicted by (2.18-2.20). (The crossover times are explicitly identified on Figure 2.4.)

Despite the reasonable match between data from the linearized and fully nonlinear systems, the linear growth assumption (2.14) for input to the central perception units, which leads to the quadratic and cubic functions of (2.17) and (2.19) is evidently too crude. In particular, the difference in inputs $i_1 - i_2$ to the decision layer departs significantly from (2.17) at large times, due to the limiting effect of the sigmoidal activation functions for large inputs, and it does not account for time delays due to the the fact that the attention units
and stimuli are co-activated, and biases to the central perception units take some time to build up. More realistic inputs can be derived by examining noise-free simulation results, as I show in Chapter 3. I also perform further reductions which replace the complicated, three-layer, eleven-unit neural network of Figure 2.2 by a DDM.
Figure 2.4: Simulations of the noise-free Eriksen model with logistic activation functions. Decision layer inputs $i_j$ (a) and outputs $z_j$ (b) for compatible stimuli, and for incompatible stimuli (c,d): crossovers in (c,d) cause the dip in accuracy. Solid blue curves indicate $i_1$ and $z_1$: correct response; dashed red curves indicate $i_2$ and $z_2$: incorrect response. Stimuli are applied at $t = 0$, after units have settled at resting values. Parameters are specified in text.
Figure 2.5: Simulations of the noise-free Eriksen model with logistic activation functions showing decision layer inputs $i_1$ and outputs $z_1$ (solid and dashed blue) and $i_2$ and $z_2$ (dash-dotted and dotted red); inputs in panels (a,c) and outputs in (b,d). Panels (a,b) show effect of attention $a_c = 0.5$ (solid and dotted) and $a_c = 1$ (dash-dotted and dotted), and panels (c,d) show effect of the ratio $b/a = 1$ (solid and dotted) and $b/a = 2$ (dashed and dash-dotted). See text for discussion. Other parameters as for Figure 2.4.
Chapter 3

A Drift Diffusion Model with Variable Drift Rate

3.1 DDM reductions

As noted in Chapter 1, during the decision process, the difference between the amounts of evidence supporting the two choices is accumulated. I assume $x(t)$ be the difference in evidence at time $t$, and assume that $x = 0$ represents equal amounts of evidence. By symmetry, $x(t) > 0$ implies that at time $t$, there is more evidence supporting the first choice, and vice versa. Also I assume that the first decision is the correct choice for this trial. I consider the unbiased condition, i.e. the initial condition is $x(0) = 0$ since there is no evidence supporting either choice. During the decision process, evidence is accumulated according to the following equation:

$$\dot{x} = A + c\eta. \hspace{1cm} (3.1)$$

In above Equation, $\dot{x}$ denotes the rate of change of $x$ over time. The right side of the equation has two parts: one is a constant drift, where $A$ denotes the average increase in evidence accumulating the correct choice per unit time; the second term $c\eta$ represents the white noise, where $c$ is a constant, and $\eta$ is white noise (samples drawn from a Gaussian
distribution with mean zero) and changes during each time step. The noise fluctuation term over each time step $dt$ is a Gaussian distribution of mean 0 and variance $c^2 dt$. Therefore, the probability density distribution $p(x, t)$ is also a Gaussian distribution with mean $At$ and standard deviation $c\sqrt{t}$, which can be written by [34]:

$$ p(x, t) = \mathcal{N}(At, c\sqrt{t}) = \frac{1}{\sqrt{2\pi c^2 t}} \exp\left(\frac{(x - At)^2}{2c^2 t}\right). \quad (3.2) $$

It has been shown that various neural network models of decision-making [21, 23, 106] can be reduced to variants of the DDM [11]. In this chapter, a DDM reduction of the neural network model of Eriksen task, which is described in Chapter 2, is derived and analyzed.

Note that the difference between the activities of the linearized decision layer units follows an Ornstein-Uhlenbeck or drift-diffusion process (cf. equations (2.6-2.7)). For convenience I rewrite the latter employing the Itô form as is common in SDE theory [4, 5]:

$$ du = [\lambda u + A(t)] dt + c dW; \quad (3.3) $$

here $\lambda = gw - k$, $A(t) = g(i_1 - i_2)$ is the time-varying drift rate, $(i_1 - i_2)$ is the net input, and $c$ denotes the r.m.s. noise strength, assumed to be constant. I discuss psychological interpretations of $A(t)$ in Section 3.3.

In an incompatible trial of the Eriksen task, the net input $(i_1 - i_2)$ is particularly interesting. Figure 3.1 (left) shows the net input $i_1 - i_2$ from a fully nonlinear but noise-free simulation of an incompatible trial, in comparison to three analytical approximations: a linear expression

$$ g(i_1 - i_2) = d_0 + d_1 t, \quad (3.4) $$

the quadratic expression (2.17), which I rewrite in the form

$$ g(i_1 - i_2) = q_0 t + q_1 t^2, \quad (3.5) $$

and the exponential

$$ g(i_1 - i_2) = a_0 + a_1 \exp(a_2 t) + a_3 \exp(a_4 t). \quad (3.6) $$
With suitable parameter choices, the former two match the nonlinear system’s response well in the middle and in the early and middle time ranges respectively, but neither captures its asymptotic approach to a constant as $t$ continues to increase (Figure 3.1). However, the exponential function (3.6) provides an excellent fit throughout. The specific parameter values obtained are:

\[
\begin{align*}
\text{linear} : g(i_1 - i_2) &= -0.258 + 0.145t, \\
\text{quadratic} : g(i_1 - i_2) &= -0.254t + 0.1420t^2, \\
\text{exponential} : g(i_1 - i_2) &= 0.476 + 6.396 \exp(-0.759t) - 6.906 \exp(-0.659t).
\end{align*}
\] (3.7)

In the compatible case stimuli reinforce rather than compete, and the attention layer again further accentuates the contribution of the center units as time progresses, leading to the monotonically increasing function shown in Figure 3.1 (right), which can also be well-fitted by an exponential, in this case requiring only three parameters:

\[
g(i_1 - i_2) = a_0 + a_1 \exp(a_2 t),
\] (3.8)

the specific values being $a_0 = 0.934$, $a_1 = -0.787$ and $a_2 = -0.960$. 

23
The number of parameters defining the exponential drift rates in the incompatible and compatible cases can each be reduced by one by requiring that the net input \( g(i_1 - i_2) = 0 \) at an appropriate reference time such as \( t = 0 \). This was not done for the fits noted above (since the quantities obtained from full simulations did not vanish at \( t = 0 \)), but I use it in fitting empirical data in §3.2.

3.1.1 The interrogation protocol

I first derive simple analytical expressions for accuracy in terms of response time and the system parameters introduced above, assuming that that responses are delivered under the interrogation protocol. Although this paradigm has not typically been used in empirical studies of performance in the Eriksen task, it provides an approximation of the deadlining procedures used to produce conditional accuracy curves of the sort shown in Figure 1.1. Such procedures are required to generate an adequate number of responses at short latencies and low accuracy.

To analyze accuracy in terms of response time under the interrogation protocol, I observe that the probability distribution \( p(u, t) \) for solutions of (3.3), derived from the associated Fokker-Planck or forward Kolmogorov equation [34]:

\[
\frac{\partial p(u, t)}{\partial t} = - \frac{\partial}{\partial u} \left[ (\lambda u + A(t)) p(u, t) \right] + \frac{c^2}{2} \frac{\partial^2 p(u, t)}{\partial u^2},
\]

may be written in the form

\[
p(u, t) = \frac{1}{\sqrt{2\pi \nu(t)}} \exp \left[ \frac{-(u - \mu(t))^2}{2\nu(t)} \right],
\]

where

\[
\mu(t) = \mu_0 e^{\lambda t} + \int_0^t e^{\lambda(t-s)} A(s) \, ds \quad \text{and} \quad \nu(t) = \nu_0 e^{2\lambda t} + \frac{c^2}{2\lambda} \left( e^{2\lambda t} - 1 \right)
\]

denote the evolving mean and variance of \( p(u, t) \), and I have assumed that initial conditions \( u(0) \) for (3.3) are drawn from a Gaussian distribution with mean \( \mu_0 \) and variance \( \sqrt{\nu_0} \). For a balanced system, \( \lambda = gw - k = 0 \) and equations (3.11) become:

\[
\mu(t) = \mu_0 + \int_0^t A(s) \, ds \quad \text{and} \quad \nu(t) = \nu_0 + c^2 t.
\]
Figure 3.2: Analytical predictions of accuracy for the drift-diffusion model under the interrogation protocol with linear (solid, blue), quadratic (dotted, red) and exponential (dashed, green) drift rates, fitted to the nonlinear simulation data of Figure 3.1 as described in the text. Parameter values for the drift-diffusion process are $\lambda = 0$ (balanced), $c = 0.3$.

I model the interrogation protocol by assuming that, on a given trial, the subject chooses the alternative that seems more probable at time $T$. Thus, response $>$ is given if $u(T) > 0$ and $<$ is given if $u(T) < 0$, and if $>$ is the correct alternative, the average fraction of correct responses (accuracy) is therefore given by:

$$P_{\text{correct}}(T) = \int_{0}^{\infty} p(u, t) \, du = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{\mu(T)}{\sqrt{2\nu(T)}} \right) \right]. \quad (3.13)$$

To make specific comparisons, I assume that $\mu_0 = \nu_0 = 0$ (initial conditions are reset to $u(0) = 0$ for each trial, corresponding to an unbiased start). I additionally take $\lambda = 0$, corresponding to the optimal drift-diffusion process [11], and consistent with the parameter choices of Chapter 2 ($w = k = 1$ and $g = 1$ for the decision layer). Substituting the linear (3.4), quadratic (2.17), and exponential (3.6) expressions into the first expression of (3.12) and appealing to (3.13), I produce the accuracy vs. mean response time curves of Figure 3.2. These all exhibit an early dip in accuracy below 50%, as in the data of [43] (cf. Figure 1.1), and I may compute the times $t_{\text{min}}$ at which their minima occur and the
crossover times $t_{50}$ at which accuracy regains 50%. The latter are given by $\mu(t_{50}) = 0$, leading to the explicit expressions

$$t_{50} = \frac{-2d_0}{d_1} \quad \text{and} \quad t_{50} = \frac{-3q_0}{2q_1}, \quad (3.14)$$

and the implicit one

$$a_0 t_{50} + \frac{a_1}{a_2} (e^{a_2 t_{50}} - 1) + \frac{a_3}{a_4} (e^{a_4 t_{50}} - 1) = 0, \quad (3.15)$$

in the linear, quadratic and exponential cases respectively. The minima occur at the (negative) turning points of the error function argument $\mu(t)/\sqrt{2\nu(t)}$, i.e.

$$t_{min} = \frac{-2d_0}{3d_1} \quad \text{and} \quad t_{min} = \frac{-9q_0}{10q_1}, \quad (3.16)$$

and the solution of

$$(a_0 + 2a_1 e^{a_2 t_{min}} + 2a_3 e^{a_4 t_{min}}) t_{min} = \frac{a_1}{a_2} (e^{a_2 t_{min}} - 1) + \frac{a_3}{a_4} (e^{a_4 t_{min}} - 1). \quad (3.17)$$

The fact that the quadratic drift rate leads to the fastest approach to 100% accuracy is due to the rapid growth of that function as $t$ increases. A linear fit, with strong negative drift at early times, leads to the lowest accuracy and latest crossover time. The exponential drift rate gives accuracies between these two cases over the early RT range, only falling below the linear drift case at larger times, as the exponential approaches its finite asymptote.

### 3.1.2 The interrogation protocol with bounded domain

While analytical solutions (3.10) are available for the unbounded DD process (3.3), this process is unrealistic in that magnitudes $|u(t)|$ can become arbitrarily large. In reality, neural firing rates and differences among them must remain bounded above and below. The original Eriksen model (2.3-2.5) respects this via the upper and lower asymptotes of its sigmoidal response functions (Figure 2.1), and I shall presently compare the above results with direct simulations of this nonlinear model, but one may also incorporate these bounds in a linear context, as I now show.
Figure 3.3: Accuracy vs. RT for the bounded diffusion process with an exponential drift rate and bounds $L$ of $0.1$ (red), $0.3$ (blue) and $0.5$ (green) bottom to top on right of figure, with noise strength $c = 0.3$. Dashed black curve shows accuracy for unbounded process. Arrows at right show theoretical accuracy limits for large RT from (3.20).

Specifically, to compute the probability distribution with bounded $|u| \leq L$, (3.9) (with $\lambda = 0$) must be solved subject to no-flux boundary conditions [119]:

$$- A(t)p(u, t) + \frac{c^2}{2} \frac{\partial p(u, t)}{\partial u} = 0 \text{ at } u = -L \text{ and } u = L. \quad (3.18)$$

This can be done analytically via separation of variables for constant drift rates $A$, but I must resort to Monte Carlo simulations in the present case of time-varying drift. Figure 3.3 shows examples to illustrate the effect of increasing the boundary $L$ from the small value of $c/3$ to the relatively large one of $5c/3$, illustrating convergence to the unbounded result of Figure 3.2 as $L$ increases. Note that small $L$ gives higher accuracy for mid-range RT responses than large $L$, but that this effect reverses for slower responses, since the asymptotic accuracy increases toward 1 as $L$ increases. This is due to the fact that more sample paths that would have remained below 0 in the unbounded case (giving errors) are reflected from the lower boundary $u = -L$ and thereafter move above 0 under the influence of positive drift, than are reflected from $+L$ and subsequently move below 0. For example, averages over 10,000 simulated trials show that for $L = 0.1$ about 24% of the sample paths reflected...
from $+L$ cross and remain below 0, while for $L = 0.3$ and $L = 0.5$ the figures are 3% and 0.3% respectively.

In the limit $t \to \infty$, the exponential drift rate becomes constant $A(t) \to A = a_0$, and at long times the probability distribution $p(u, t)$ approaches the equilibrium solution of (3.9) with $\lambda = 0$ and $A(t) \equiv A$ and boundary conditions (3.18):

$$p(u, t) \to p_{eq}(u) = \frac{2A \exp \left( \frac{2Au}{c^2} \right)}{c^2 \left[ \exp \left( \frac{2AL}{c^2} \right) - \exp \left( \frac{-2AL}{c^2} \right) \right]}.$$  (3.19)

The asymptotic accuracy is therefore

$$\lim_{t \to \infty} P_{\text{correct}}(t) = \int_0^L p_{eq}(u) du = \frac{1}{1 + \exp \left( \frac{-2AL}{c^2} \right)} < 1.$$  (3.20)

Now $P_{\text{correct}}(t) \to 1$ as $L \to \infty$, but asymptotic accuracy is bounded for all finite $L$. The analytical $t \to \infty$ limits are indicated for three values of $L$ in Figure 3.3. In contrast, for the linear and quadratic drift rates of (3.4-3.5) that both grow without bound, accuracies approach 100% as $t$ increases (not shown).

### 3.1.3 The free-response protocol

For the free-response protocol the appropriate setting is that of a first passage problem, as treated in considerable detail for both DDM and OU processes in [11, Appendix A], cf. [34]. Unfortunately, to my knowledge, neither explicit solutions nor asymptotic approximations are available for first passage problems with time-varying drift rates, and so I must again use numerical methods. In Figure 3.4 (left panel) I compare the analytical interrogation results of Section 3.1.1 with simulations of the full Eriksen model, and in the right panel I compare free response data from the Eriksen model with numerical simulations of the DDM with exponential drift rate inputs as fitted in Figure 3.1, Eqns. (3.8-3.7). Parameter values for the results here and in Figure 3.5 were: $k = w = l = h = 1$, $a = b = a_c = 1$ and $g = 0.55$, $b = 0.8$ for the attention and perception units and $g = 1$, $b = -0.9$ for the decision units. With the exception of the stimulus strengths $a$ and $b$, these are identical to those used for my earlier simulations and lead to a pure DDM with $\lambda = 0$. Threshold values are given in the figure captions.
Figure 3.4: Left: Accuracy vs. RT for the full Eriksen model under the interrogation protocol (solid line) compared with analytical results for the unbounded DDM (stars). Right: Accuracy vs. RT for the full Eriksen model (solid line) under the free-response protocol compared with numerical results for the unbounded DDM with exponential inputs (stars): thresholds are $\theta = 1.0$ for full model and $\theta = \pm 0.3$ for DDM. Upper curves denote compatible trials and lower curves incompatible trials. Remaining parameters are as specified in text.

The full model gives higher accuracy at early times and lower accuracy at later times than the unbounded DDM for incompatible stimuli in the interrogation case, much as does the bounded DDM of Figure 3.3, suggesting that the limiting nature of the sigmoidal response function, ignored in the unbounded linearized analysis, comes into play. The numerical free-response results from the linearized DDM exhibit leftward (time compression) shifts in both compatible and incompatible cases compared with the full simulations. The analytical interrogation results are closer to the full nonlinear simulations for incompatible trials, but a similar shift leads to overestimates of accuracy at early response times.

The free response accuracy results are rather sensitive to the choice of threshold in the DDM, which I based on the difference between the noise-free steady states of the decision layer outputs, as illustrated in Figures 2.4-2.5 (note that those figures were computed for parameters differing from those given directly above). Specifically, noting that the DDM net evidence variable is $u = z_1 - z_2$ (Equations (2.6-2.7)), it follows that if $z_1$ and $z_2$ equilibrate
Figure 3.5: Accuracy vs. RT for the full Eriksen model (blue lines with star dots) under the free-response protocol, with RT histograms below (left, dark blue bars of each pair). Unbounded DDM accuracy results are shown for comparison in solid red with open triangles and RT histograms in right, red bars of each pair. Thresholds are $\theta = 1.1$ for full model and $\theta = \pm 0.3$ for DDM. Left panel shows compatible trials, right panel shows incompatible trials. Remaining parameters are as specified in text.

to steady state levels $z_{1}^{\infty} > z_{2}^{\infty}$ as $t \to \infty$ in the noise-free simulation with stimulus 1 ($<$) applied, the appropriate threshold for $u$ must lie in the range $(0, z_{1}^{\infty} - z_{2}^{\infty})$. Adjustments to match accuracy curves may be made in this range. Also due to this transformation from $z$ to $u$, the results of coordinate rotation change the threshold value $\theta$. Details of how to transform thresholds can be found in [11] Figure 7. Figure 3.5 shows the data from the full Eriksen model again, with DDM data obtained using a modified threshold that provides a better match. I also show reaction-time histograms from the full simulation and from the DDM. Note that, as in the experiments, reaction times are longer under the incompatible condition than the compatible condition, and that the DDM provides reasonable estimates of RT distributions, especially in the incompatible case.
3.2 Comparisons with empirical data

In this section I compare results from the Eriksen task model of §2.2 with the pure $\lambda = 0$ DDM of §3.1, working under the free response (threshold crossing) paradigm, with the data of [43]. I reproduce the model fit described in [23], and perform a new fit to determine values of the parameters $a_j$ describing the exponential drift rates for both compatible and incompatible stimuli, as in Chapter 2 §2.4. Figures 3.6 and 3.7 show the resulting accuracy curves and RT distributions in comparison with the experimental data (cf. Figure 1.1 above). A visual inspection of Figure 3.6 shows that the DDM fits the accuracy data somewhat better than the full nonlinear model. I quantify and comment further on this below.

Data fits for the DDM were performed using the \texttt{fmincon()} function in MATLAB for comparison with the fits to the original connectionist model of [23]. As the analyses of §2.4 and §3.1 predict, different exponentially-varying drift rates $A(t)$ are required for incompatible and compatible cases: the former having 5 and the latter 3 parameters (Equations (3.6) and (3.8)). To reduce the number of free parameters, it is required that $A(0) = 0$, thereby reducing these numbers to 4 and 2 respectively. The noise variance $c$ and threshold $\theta$ add 2
more parameters, for a total of 8. These 8 parameters were determined by adjusting them while seeking minima of a fitting error function which averages over all the accuracy and reaction time data for compatible and incompatible trials.

The fitting error utilizes a weighted Euclidean norm. The usual Euclidean ($L^2$) distance between vectors $\mathbf{u}$ and $\mathbf{v}$ with components $u_j$ and $v_j$ is

$$||\mathbf{u} - \mathbf{v}|| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \ldots + (u_n - v_n)^2}.$$  

(3.21)

Figure 3.7: RT histograms from the full model (top row) and the DDM (bottom row), compared with data from [43]. Model results are shown in right, yellow bars and empirical data in left, blue bars for each 50 ms bin. Compatible trials appear in left column and incompatible trials in right column.
by (3.21). Since the units of accuracy and reaction time differ, each of these was then weighted by dividing it by the mean of the data, indicated by an overbar. This produces the non-dimensional quantity:

$$\text{Error} = \sum_{\text{com.incom}} \left[ \frac{||AC_d - AC_m||}{||AC_d||} + \frac{||RT_d - RT_m||}{||RT_d||} \right].$$

(3.22)

This represents the sum of the percentage differences in accuracy and reaction time.

According to the error measure of (3.22), the DDM provides a 24% improvement over the fit obtained for the full connectionist model of [23]. Moreover, this is achieved using 8 parameters in comparison with 11 for the connectionist model.

### 3.3 Discussion

In Chapters 2 and 3 of this dissertation I analyze a linearized version of the connectionist model for the Eriksen two-alternative forced-choice flanker task presented in [23] and [92]. I show that, provided solutions remain within the central domain of the logistic function in which it may be approximated by a linear function that matches its slope at the bias point $\beta$, as proposed by [21], analytical solutions of a decoupled, linearized model modulated by a pre-determined attention signal can provide reasonable estimates of critical times at which evidence in favor of the correct and incorrect alternatives cross over for incompatible trials and hence reproduce the characteristic dip in accuracy for such trials. I also show that the dynamics of the two-unit decision layer can be decoupled and reduced to a drift-diffusion model (DDM) whose drift rate represents the net evidence for one alternative coming from the perception layer.

I then derive estimates of accuracy as a function of response time by interrogating a DDM with variable drift rates that are fitted to outputs from the perception layer of the fully nonlinear model. Collapsing to this model reduces the number of parameters from 11 or more in the connectionist model to 8 in the DDM with exponential drift rates. I compute the evolving probability distribution of solutions to the DDM and integrate it to obtain the psychometric function (% correct) as an explicit function of response time.
and the parameters defining the drift rate and noise strength. The interrogation protocol assumes that the response delivered reflects the subject’s current estimate, and corresponds best to a deadline task with a cued response.

I also consider a protocol under which subjects respond in their own time, modeled as a first passage problem. The qualitative forms of psychometric functions in the interrogation and free response cases are similar to those of the full nonlinear model for both compatible and incompatible trials, the latter showing the characteristic dip below chance for early responses. The DDM also produces acceptable approximations to accuracy and reaction time distributions derived from simulations of the full nonlinear model, and, more strikingly, it provides a slightly better fit to empirical data than does the full model, while using fewer parameters.

These results show that judicious linearization and decoupling of processing layers in connectionist models can allow analytical studies of how parameters influence the behavior of such models. They also suggest that parameter tuning based on the explicit formula available for the DDM interrogation protocol may be generally useful in matching model results with behavioral data, even under free response protocols. The key linearization step has been justified in model studies of 2AFC tasks [17], and extended to multiple alternative decision models [71]. The range over which response functions $\psi$ are well-approximated by their linearizations grows with the dynamic range of the neurons involved (the output range is normalized to 1 in Eqn. (2.2)). Decoupling is more problematic: a simple a priori assumption regarding biases due to attention does not produce realistic inputs from the perception layer to the decision layer or to the DDM, although such inputs can be derived from simulations of the full network, and here they are accurately fitted by simple exponential functions.

The present study also provides a foundation for further theoretical and experimental work. The DDM reduction reveals that interaction of the attention and perception layers produces variable drift rates, implying varying signal-to-noise ratios. These modulate the conjectured integration of evidence (in LIP) as attention is progressively engaged by top-
down control. This interpretation is consistent with the assumption, within the DDM framework, that attention modulates drift rate.

These findings present an interesting challenge to the hypothesis that human performance in 2AFC tasks reflects the operation of optimal decision making processes. In a stationary environment such as the one modeled here, a constant drift-diffusion process produces optimal behavior [11]. This contrasts with my observation of variable drift, suggesting that human performance in the Eriksen flanker task is not optimal. Such sub-optimality may reflect adaptive biases developed in response to a broader class of experience. For example, a Bayesian approach to analyzing optimality in this task [117] has shown that prior expectations of compatible stimuli, or of correlations between neighboring elements in the visual field, can produce the observed dips in accuracy. Hence, experience of natural environments may have biased perception and decision systems to expect response-compatibility or perceptual similarity of nearby inputs. Work relating such expectations to the dynamics of processing indicate that such biases can be related directly to the variable drift rates incorporated in the present model, and used to account for empirical observations concerning performance as described in the next chapter.

In summary, the work reported here exemplifies how connectionist models of the dynamics of processing in cognitive tasks can be related to the DDM, which then provides an analytic framework for interpreting empirical observations. Specifically, I have shown that a DDM reduction of a multi-layer model of the Eriksen task suggests that processing involves a progressive change in the drift rate over the course of a trial, reflecting the influence of top-down attentional mechanisms. This finding can be related, in turn, to an optimality analysis that generates new hypotheses about the factors governing attentional control (e.g., prior expectations of stimulus compatibility). The present work therefore provides links between theoretical analysis of optimal performance and formal specifications of the dynamics of neural processing mechanisms responsible for actual performance.
Chapter 4

Dynamical Systems Analysis of Inference Models for the Eriksen Task

4.1 Bayesian inference models for neural systems

Bayesian inference is statistical inference in which incoming evidence is used to update or to newly infer the probability that a hypothesis may be true. The name “Bayesian” comes from the frequent use of Bayes’ theorem in the inference process [8]. Bayesian inference usually relies on subjective probabilities, in the induction process. Bayesian statisticians believe that probabilities can also have an objective value and therefore Bayesian inference can provide an objective method of induction.

Bayesian inference techniques have been a fundamental part of computerized pattern recognition techniques since the late 1950s. There is also an ever growing connection between Bayesian methods and Monte Carlo simulation techniques. More recently, theoretical work based on Bayesian probability theory has suggested that under certain task conditions, differential processing is optimal and thus desirable, independent of any resource limitation considerations [26, 25]. On the other hand, experimental data have also shown that, under
specific experimental settings such as multi-modal integration based on noisy cues, human subjects can indeed integrate differentially reliable sensory cues in a Bayes-optimal fashion [31, 62].

Yu et al. [117] provide a Bayesian framework for dynamic attentional selection and illustrate it using the Eriksen flanker task. They show how two distinct Bayesian inferential principles can explain the experimental findings. They also suggest ways in which these theoretical results can be tested experimentally, as well as their more general significance for an understanding of the mechanisms underlying attention. In this Chapter, I analyze the dynamics of the Bayesian models, which are nonlinear, coupled discrete-time dynamical systems, by considering simplified, approximate systems that are linear and decoupled. Analytical solutions of these allow a description of how posterior probabilities and psychometric functions depend upon model parameters. I investigate the continuum limits of these simplified dynamical systems, and demonstrate that Bayesian updating is closely related to a drift-diffusion process, whose implementation in neural network models has been extensively studied (e.g. see Chapters 2-3 of this dissertation). This provides insight on how neural substrates can implement Bayesian computations. I also compare my results with numerical simulations of the original models and show that the agreement is rather good. Finally, I provide direct comparisons between behavioral data and predictions of the full and approximate compatibility bias models.

4.2 A Bayesian framework for the Eriksen task

I start by briefly reviewing the compatibility bias and spatial uncertainty inference models for the Eriksen task proposed by Yu et al. [117]. The generative process common to both models consists of the prior probability distribution over trial type ($M = 1$ if compatible, $M = 2$ if incompatible), and the stochastic relationship between the trial type $M$ and the stimuli $s$, and between the stimuli and the noisy inputs into the visual system $x$. For simplicity, it is assumed that there are three stimuli, $s \triangleq \{s_1, s_2, s_3\}$, for “left”, “center”, “right”, respectively; and that each one of three neural units or populations $x \triangleq \{x_1, x_2, x_3\}$
responds to one stimulus. Here the pairs of left and right flankers are combined in $s_1$ and $s_3$ respectively, and assume that all three inputs contain independent noise, both among the units/populations, and over time. Using integers $s_i = \pm 1$ to denote < and >, and $M = 1, 2$ to denote compatible and incompatible trials respectively, the process may be formally described as:

$$\beta \triangleq P(M = 1) \in [0, 1]$$  \hspace{1cm} (4.1)

$$P(s|M = 1) = \begin{cases} 0.5 & s_1 = s_2 = s_3 = -1 \quad (<<<<<) \\ 0.5 & s_1 = s_2 = s_3 = +1 \quad (>>>>>) \end{cases}$$  \hspace{1cm} (4.2)

$$P(s|M = 2) = \begin{cases} 0.5 & s_1 = s_3 = +1, s_2 = -1 \quad (<<>>>) \\ 0.5 & s_1 = s_3 = -1, s_2 = +1 \quad (>>>>) \end{cases}$$  \hspace{1cm} (4.3)

$$p(x_i|s) = p(x_1(t)|s)p(x_2(t)|s)p(x_3(t)|s),$$  \hspace{1cm} (4.4)

$$p(x_1, x_2, \ldots, x_t|s) = p(x_1|s)p(x_2|s)\ldots p(x_t|s).$$  \hspace{1cm} (4.5)

For the compatibility bias model, the prior probability $\beta$ for compatible trials is assumed to be higher than the “true” value 0.5, and the inputs are taken to be normally distributed as a function of their respective stimuli, and independent of neighboring stimuli:

$$p(x_i|s) = p(x_i(t)|s_i) = \frac{1}{\sqrt{2\pi}\sigma_i^2} \exp \left[ -\frac{(x_i - s_i)^2}{2\sigma_i^2} \right].$$  \hspace{1cm} (4.6)

Thus, at each step $t$ the $x_i(t)$ are independently drawn from normal distributions with means $s_i$ and standard deviations $\sigma_i$. I denote this procedure below by $x_i(t) \sim \mathcal{N}(s_i, \sigma_i^2)$.

In the spatial uncertainty model, the correct prior $\beta = 0.5$ is assumed, but the inputs are corrupted by their neighbors according to:

$$p(x_i|s) = p(x_1(t)|s_1, s_2) p(x_2(t)|s_2, s_3) p(x_3(t)|s_2, s_3),$$  \hspace{1cm} (4.7)

where

$$x_1(t) \sim \mathcal{N}(a_1s_1 + a_2s_2, \sigma_1^2 + \sigma_2^2),$$

$$x_2(t) \sim \mathcal{N}(a_1s_2 + a_2s_1 + a_2s_3, \sigma_1^2 + 2\sigma_2^2),$$

$$x_3(t) \sim \mathcal{N}(a_1s_3 + a_2s_2, \sigma_1^2 + \sigma_2^2).$$  \hspace{1cm} (4.8)
where $a_1$, $\sigma_1$ denote influence from the primary stimulus, and $a_2$ and $\sigma_2$ that from the flankers.

Define $z_{i,j}^t \triangleq P(s_2 = i, M = j|X_t)$ for the posterior probabilities, and $l_{i,j}^t \triangleq p(x_t|s_2 = i, M = j)$ for the likelihood functions, where $i \in \{-1, +1\}$, $j \in \{1, 2\}$. Based on Bayes’ Rule, this yields the inference model: four discrete-time dynamical equations, coupled through normalization:

$$z_{i,j}^t = \frac{l_{i,j}^t z_{i,j}^{t-1}}{\sum_{k,l} l_{k,l}^t z_{k,l}^{t-1}},$$

with initial conditions

$$z_{i,j}^0 = \begin{cases} \beta^2, & j = 1, \forall i; \\ \frac{(1-\beta)^2}{2}, & j = 2, \forall i. \end{cases}$$

To make a decision based on the accumulating inputs, compare the cumulative marginal posterior probability,

$$P(s_2 = i|X_t) = z_{i,1}^t + z_{i,2}^t,$$

against a decision threshold $q$, a policy closely related to the SPRT [108]. As soon as $P(s_2 = i|X_t)$ exceeds $q$ for $i = -1$ or $i = +1$, the system chooses the corresponding response (H or S) and terminates observations for the current trial. The computation for the marginal posterior probability over compatibility is analogous: $P(M = j|X_t) = z_{-1,j}^t + z_{1,j}^t$.

For specific calculations, unless otherwise noted, the parameter values used in [117] are adopted, namely: $\sigma = 9$ for the compatibility bias model and $\sigma_1 = 7$, $\sigma_2 = 5$, $a_1 = 0.7$, $a_2 = 0.3$ for the spatial uncertainty model and $q = 0.9$ for both.

### 4.3 Linearization and parametric dependence

It was shown in [117] that certain choices of parameters allow both the compatibility bias and spatial uncertainty models to capture key properties of the behavioral data in Fig. 1.1 (see Figures 4.9-4.10 below). Here I derive general constraints on the parameters in each model that allow them to reproduce the behavioral data: $\sigma$ for the compatibility bias model, $a_1$, $a_2$, $\sigma_1$, and $\sigma_2$ for the spatial uncertainty model; and $n$, the number of distractors. While I
cannot analyze the complex relationship between accuracy and reaction time directly, I wish

to at least constrain parameters so that the mean posterior probability for \( s_2 = 1 \) (the correct
answer) dips below that for \( s_2 = -1 \) after one or a few time steps of observations. Although

the relative probability of a correct response at time \( t \) depends not just on the mean but also

on higher-order moments, such an analysis would illuminate the magnitude and range of

the effective parameters. Unfortunately, even this partial analysis is difficult for the original

Bayesian model, since \( P(s_2|X_t) \) involves the summation of two exponential functions of the

inputs, as in Eq. (4.11), and there is no obvious way to derive the expectation of \( P(s_2|X_t) \)

as an explicit function of the parameters that specify the generation of the inputs \( x \).

Due to such computational intractability, I instead work with a linearized approximation

to the exact posterior update rule of Eq. (4.9). I will motivate and describe the approxi-
mations for the two Bayesian models, and demonstrate via simulations that the parametric

constraints derived from this approximate scheme provide useful bounds for the original

Bayesian models.

4.3.1 The compatibility bias model

Given the assumption of independent, normally-distributed inputs (Eqs. (4.4) and (4.6)), I

have

\[
p(x_t|s) = \left\{ \frac{\exp\left[\frac{- (x_1 - s_1)^2}{2 \sigma^2}\right]}{\sqrt{2 \pi \sigma^2}} \right\} \left\{ \frac{\exp\left[\frac{- (x_2 - s_2)^2}{2 \sigma^2}\right]}{\sqrt{2 \pi \sigma^2}} \right\} \left\{ \frac{\exp\left[\frac{- (x_3 - s_3)^2}{2 \sigma^2}\right]}{\sqrt{2 \pi \sigma^2}} \right\},
\]

where each \( s_i \) can take on the value \( \pm 1 \). I now derive an approximation to Eq. (4.12) that

is linear in the \( x_i(t) \)'s. Defining

\[
\Theta_k \equiv \frac{\exp\left[\frac{- (x_k - 1)^2}{2 \sigma^2}\right]}{\sqrt{2 \pi \sigma^2}} \quad + \quad \frac{\exp\left[\frac{- (x_k + 1)^2}{2 \sigma^2}\right]}{\sqrt{2 \pi \sigma^2}},
\]

40
the likelihood function for \( s_2 = 1 \) and \( M = 1 \) (i.e. \( s_1 = s_2 = s_3 = 1 \)) can be approximated as follows:

\[
p(x_t | s) = \Theta_1 \Theta_2 \Theta_3 \left\{ \frac{\exp \left[ -\frac{(x_1-1)^2}{2\sigma^2} \right]}{\Theta_1 \sqrt{2\pi\sigma^2}} \frac{\exp \left[ -\frac{(x_2-1)^2}{2\sigma^2} \right]}{\Theta_2 \sqrt{2\pi\sigma^2}} \frac{\exp \left[ -\frac{(x_3-1)^2}{2\sigma^2} \right]}{\Theta_3 \sqrt{2\pi\sigma^2}} \right\}
\]

\[
= \Theta_1 \Theta_2 \Theta_3 \frac{1}{1 + \exp \left( -\frac{2x_1}{\sigma^2} \right) \frac{1}{1 + \exp \left(-\frac{2x_2}{\sigma^2}\right) \frac{1}{1 + \exp \left(-\frac{2x_3}{\sigma^2}\right)}}
\]

\[
= \frac{\Theta_1 \Theta_2 \Theta_3}{8} \left[ 1 + \frac{x_1}{\sigma^2} \right] \left[ 1 + \frac{x_2}{\sigma^2} \right] \left[ 1 + \frac{x_3}{\sigma^2} \right] + \mathcal{O}(x_k^2/\sigma^4)
\]

\[
\approx \frac{\Theta_1 \Theta_2 \Theta_3}{8} \left[ 1 + \frac{x_1 + x_2 + x_3}{\sigma^2} \right] + \mathcal{O}(x_k^2/\sigma^4)
\]

(4.13)

The first step uses the fact that quadratic and constant terms cancel in the ratios, the next two rely on Taylor series expansion of the exponential terms and the binomial series approximation:

\[
\left[ 1 + \exp \left( -\frac{2x_k}{\sigma^2} \right) \right]^{-1} \approx \left[ 2 \left( 1 - \frac{x_k}{\sigma^2} \right) \right]^{-1} \approx \frac{1}{2} \left[ 1 + \frac{x_k}{\sigma^2} \right]
\]

and the approximation is justified by the fact that \( x_k(t) \in [-1 - 2\sigma, 1 + 2\sigma] \) with > 99% probability, if I can assume that \( \sigma \gg 1 \). This latter assumption is reasonable since I am modeling the time-scale at which on average many time steps of inputs are needed to perform the discrimination.

Generalizing the approximation (4.13) to the other three cases and using the four resulting expressions in Eq. (4.9), yields the following approximate update rules:

\[
z_{i,j}^t \approx \frac{1}{D_t} \left\{ \begin{array}{ll}
\left[ 1 + \frac{x_1(t)+x_2(t)+x_3(t)}{\sigma^2} \right] z_{+1,1}^{t-1} & s_2 = +1, M = 1 , \\
\left[ 1 - \frac{x_1(t)+x_2(t)+x_3(t)}{\sigma^2} \right] z_{-1,1}^{t-1} & s_2 = -1, M = 1 , \\
\left[ 1 + \frac{x_1(t)-x_2(t)+x_3(t)}{\sigma^2} \right] z_{+1,2}^{t-1} & s_2 = +1, M = 2 , \\
\left[ 1 - \frac{x_1(t)-x_2(t)+x_3(t)}{\sigma^2} \right] z_{-1,2}^{t-1} & s_2 = -1, M = 2 , 
\end{array} \right.
\]

(4.14)

where the denominator \( D_t \) is the sum of all four numerators and normalizes the posterior distribution, and the common factors \( \Theta_1 \Theta_2 \Theta_3/8 \) in the numerators and denominator of
Eq. (4.14) have canceled. Initial conditions are as in Eq. (4.10). Since this simplified system is still nonlinearly coupled through the denominator $D_t$, work with the joint probability $v_{i,j}^t \equiv p(s_2 = i, M = j, X_t)$ instead. The two are related as follows:

$$z_{i,j}^t = P(s_2 = i, M = j|X_t) = \frac{P(s_2 = i, M = j|X_{t-1})p(x_t|s_2 = i, M = j)}{p(x_t|X_{t-1})} = \frac{P(s_2 = i, M = j) \prod_{t'=1}^{t-1} p(x_{t'}|s_2 = i, M = j)}{p(X_t)} = \frac{v_{i,j}^t}{\sum_{k,l} v_{k,l}^t}. \quad (4.15)$$

The joint probability $v_{i,j}^t$ obeys the uncoupled update rule:

$$v_{i,j}^t = v_{i,j}^t v_{i,j}^{t-1} \approx \left(1 + \frac{\pm x_1(t) \pm x_2(t) \pm x_3(t)}{\sigma^2}\right) v_{i,j}^{t-1}, \quad (4.16)$$

where the sign preceding each $x_i$ depends on $i$ and $j$ as in Eq. (4.14). As is apparent in Eq. (4.15), $z_{i,j}^t$ can be obtained by normalizing $v_{i,j}^t$ on time step $t$, but $v_{i,j}^t$ cannot be used directly in the perceptual decision process, since a fixed threshold in the posterior probability space has no equivalent fixed value in the joint posterior space. However, $v_{i,j}^t$ is sufficient for deriving bounds on the generative parameters that on average make the posterior probability for $s_2 = 1$ dip below that for $s_2 = -1$ after one time step, when the inputs are generated from the incompatible stimulus array: $s = (-1, 1, -1)$ (the analysis for $s = (1, -1, 1)$ is analogous). Specifically, since $P(s_2, M|X_t) = p(s_2, M, X_t)/p(X_t)$, the condition $\langle z_{1,1}^t + z_{1,2}^t \rangle < \langle z_{-1,1}^t + z_{-1,2}^t \rangle$ is equivalent to $\langle v_{1,1}^t \rangle + \langle v_{1,2}^t \rangle < \langle v_{-1,1}^t \rangle + \langle v_{-1,2}^t \rangle$. Therefore require

$$\beta \left(1 - \frac{1}{\sigma^2}\right) + (1 - \beta) \left(1 + \frac{3}{\sigma^2}\right) < \beta \left(1 + \frac{1}{\sigma^2}\right) + (1 - \beta) \left(1 - \frac{3}{\sigma^2}\right) \Rightarrow \beta > \frac{3}{4}, \quad (4.17)$$

since the mean values of $x_1$, $x_2$, and $x_3$ are $-1$, $1$, and $-1$, respectively, and the compatible terms are weighted by the compatibility prior bias $\beta$ (and the incompatible ones weighted by $1 - \beta$).

Hence $\beta > 3/4$ is the necessary and sufficient condition for the average posterior probability for $s_2 = 1$ to dip below that for $s_2 = -1$ after one observation, when the true stimuli are the incompatible array $(-1, 1, -1)$. More generally, it can be shown that the constraint is $\beta > (n + 1)/(2n)$, where $n$ is the total number of flankers. This makes intuitive sense,
for it suggests that the dip is more prominent or more likely to occur when the subject’s prior compatibility bias is stronger and/or the number of flankers is larger. Indeed, there is behavioral data suggesting that flanker interference effects are stronger when there is as a lower frequency of incompatible trials [44].

Figure 4.1: Simulated and analytical approximations of parameter values that produce dips in accuracy vs. reaction time for incompatible trials. Graphs show accuracy averaged over trials with simulated reaction times under 20 time steps, as a function of $\beta$ for the compatibility bias model (left), and the ratio of means $a_1/a_2$ for the spatial uncertainly model (right). Crossings with the 0.5 accuracy line indicate numerically obtained estimates of the “true” parameter constraints; dashed vertical lines show the approximate constraints of Eqs. (4.17) and (4.23).

These analytical constraints only guarantee a dip in the posterior probability. As shown in Figure 4.1 (left), for a particular set of model parameters, the mean accuracy in compatible trials terminating within 20 time steps steadily decreases as a function of $\beta$ and drops below .5, indicating the presence of a dip, for all values of $\beta > 0.82$: somewhat higher than $\beta = 0.75$, the lower bound of inequality (4.17) that results in a dip in posterior probability. A major factor underlying the discrepancy between the two constraints is that only the mean of the posterior probability is considered, and not the full distribution. The mean accuracy depends not only on the mean posterior value, but also on higher moments. If
the distribution were symmetric about its mean, then the dip in the mean posterior would
directly translate into a dip in accuracy, but as I will show in Section 4.4, the distribution
of the posterior trajectories is strongly skewed, and the interaction of that skewness with
the decision threshold also plays a role in determining the presence of the dip in accuracy
versus reaction time.

A second reason for the discrepancy is that the theoretical bounds are for the dip to
occur in the posterior after one time step, whereas in the numerical simulations, due to
the infrequency of responses at very short RT’s, all trials that terminate within the first 20
time steps were used to estimate accuracy. If the temporal extent of the dip in the posterior
distribution is very small (which is likely in boundary cases), then conditional accuracy may
not fall below 0.5 when averaged over 20 time steps. The numerically-obtained constraints
are therefore likely to be more conservative than the analytical approximations.

### 4.3.2 The spatial uncertainty model

Derivation of iterated maps for the spatial uncertainty model are more tedious than those
of (4.14) due to the extra “cross-talk” links in the generative model, but they follow from
similar reasoning. Defining \( h_{k,i,j}^t \triangleq p(x_k(t)|s_k = i, M = j) \), forming the triple product and
dividing through by

\[
\Theta' = \prod_{k=1}^{3} [h_{k,1,1}^t + h_{k,-1,1}^t + h_{k,1,2}^t + h_{k,-1,2}^t],
\]

I obtain the approximate update rule:

\[
z_{i,j}^t = \frac{1}{D_{t-1}} \times \left\{ \begin{array}{l}
[1 - A_1 + (A_2 + A_3)(x_1 + x_3) + (A_4 + A_5)x_2] z_{1,1}^{t-1}, \\
[1 - A_1 - (A_2 + A_3)(x_1 + x_3) - (A_4 + A_5)x_2] z_{-1,1}^{t-1}, \\
[1 + A_1 - (A_2 - A_3)(x_1 + x_3) + (A_4 - A_5)x_2] z_{1,2}^{t-1}, \\
[1 + A_1 + (A_2 - A_3)(x_1 + x_3) - (A_4 - A_5)x_2] z_{-1,2}^{t-1}, \\
\end{array} \right. 
\]

(4.19)
where $D'_t$ is again the sum of the numerators and the parameters $A_i$ are

\[
A_1 = 2a_1a_2 \left( \frac{1}{\sigma_1^2 + \sigma_2^2} + \frac{1}{\sigma_1^2 + 2\sigma_2^2} \right),
A_2 = \frac{a_1}{\sigma_1^2 + \sigma_2^2},
A_3 = \frac{a_2}{\sigma_1^2 + \sigma_2^2},
A_4 = \frac{a_1}{\sigma_1^2 + 2\sigma_2^2},
A_5 = \frac{2a_2}{\sigma_1^2 + 2\sigma_2^2}.
\]  

(4.20)

Since the prior distribution is uniform, the initial conditions for (4.19) are

\[z_{i,j}^0 = \frac{1}{4}, \text{ for } i = \pm 1 \text{ and } j = 1 \text{ or } 2. \]  

(4.21)

Again, the constraint

\[
\langle P(s_2=1,M=1|X_1) \rangle + \langle P(s_2=1,M=2|X_1) \rangle < \langle P(s_2=-1,M=1|X_1) \rangle + \langle P(s_2=-1,M=2|X_1) \rangle .
\]  

(4.22)

is satisfied if $A_4(a_1 - 2a_2) < 2A_5(a_1 - a_2)$, or equivalently, if the ratio of means, $a_1/a_2$, lies in the interval

\[
\left[ \frac{2r + 3 - \sqrt{2r^2 + 6r + 5}}{1 + r}, \frac{2r + 3 + \sqrt{2r^2 + 6r + 5}}{1 + r} \right],
\]  

(4.23)

where $r \triangleq \sigma_1^2/\sigma_2^2$ is the ratio of the variances. Intuitively, if the ratio $a_1/a_2$ is too large, the flankers have negligible effects; if it is too small, the inputs lose their spatial selectivity altogether. More generally, if there are $n$ flankers, the range is the interval between

\[
\left( \frac{n}{2} + 1 \right) r + n + 1 - \sqrt{\left( \frac{n^2}{4} + 1 \right) r^2 + (n^2 + 2) r + n^2 + 1} \quad \frac{1}{1 + r}
\]  

(4.24)

and

\[
\left( \frac{n}{2} + 1 \right) r + n + 1 + \sqrt{\left( \frac{n^2}{4} + 1 \right) r^2 + (n^2 + 2) r + n^2 + 1} \quad \frac{1}{1 + r}
\]  

(4.25)

I now compare these constraints with numerical simulations of the full inference model for the specific noise parameters ($\sigma_1 = 7, \sigma_2 = 5$). The full model was simulated using a range of values of $a_1$ and $a_2$ (with their sum held at 1), and accuracy of all responses falling within the first 20 time steps was obtained as a function of $a_1/a_2$. As can be seen in Figure 4.1 (right), the accuracy in this short-RT bin is less than 0.5 when $a_1/a_2$ falls within $(0.70, 3.55)$, a somewhat more stringent condition than the analytically derived (approximate) interval $(0.67, 3.98)$. 45
4.4 A continuum limit

The key difficulty in working with the discrete dynamical systems (4.14) and (4.19) lies in the nonlinear coupling of the posteriors $z_{i,j}^t$ through the denominators $D_t$ and $D'_t$. It can be proved that individual sample paths generated with the same noise inputs are identical whether computed by iteration of Eqs. (4.14) and (4.19) or by the analogous uncoupled systems Eq. (4.16), with posteriors normalized only at the last time step; cf. Eq. (4.15). (In computing the values for the approximate model (4.16) at each step $t$ for Figure 4.8, normalization was applied only at that step, but not at steps $1$ through $t-1$, while the full iteration (4.9) is normalized at every step.) However, it does not follow that I may average over many realizations of the unnormalized process, and then normalize: as discussed further in Section 4.4.3, these operations do not commute. Nonetheless, I can decouple the dynamics by replacing the normalization constant $D_t$ at each time step with its expectation $\langle D_t \rangle$, which does not depend on the inputs, and replacing that in turn by a constant. I then take continuum limits of the resulting decoupled linear systems to form stochastic differential equations (SDEs), allowing us to use simple analytical results to compute properties of interest. As described further in section 4.6 and Chapter 6, these SDEs may in turn be related to neurally-based models of evidence accumulation.

4.4.1 Approximating the denominators

First examine the denominator $\langle D_t \rangle$ for the compatibility bias model:

$$\langle D_t \rangle = \left\langle 1 + \frac{x_1(t) + x_2(t) + x_3(t)}{\sigma^2} \right\rangle \langle z_{+1,1}^{t-1} \rangle + \left\langle 1 - \frac{x_1(t) + x_2(t) + x_3(t)}{\sigma^2} \right\rangle \langle z_{-1,1}^{t-1} \rangle$$

$$+ \left\langle 1 - \frac{x_1(t) - x_2(t) + x_3(t)}{\sigma^2} \right\rangle \langle z_{+1,2}^{t-1} \rangle + \left\langle 1 + \frac{x_1(t) - x_2(t) + x_3(t)}{\sigma^2} \right\rangle \langle z_{-1,2}^{t-1} \rangle,$$

where the approximation comes from assuming that the input-dependent terms (functions of $x_k(t)$) are independent of the $z_{ij}$ terms, which depend on the previous inputs $x_k(1),...,x_k(t)$. Although the inputs are conditionally independent (cf. Eq. (4.5)), they are marginally dependent. That is, if previous inputs favored a particular setting of $s_2$ and $M$, then the current one also tends to do the same. For analytical simplicity, I ignore this statistical
dependence. Note that in the limit as \( t \to \infty \), one of the \( z_{i,j}^t \)'s (corresponding to the actual stimulus setting) converges to 1 (and the others to 0), and that no matter which \( z_{i,j}^t \) it is,

\[
\langle D_t \rangle \to 1 + \frac{3}{\sigma^2}.
\]  

(4.26)

More generally, I expect \( \langle D_t \rangle \) to increase from 1 (\( D_0 \) is just the sum of the priors) to \( 1 + \frac{3\mu}{\sigma^2} \), where \( \mu \) denotes the mean value of the \( x_j \)'s. Figure 4.2 confirms exactly this for both compatible and incompatible stimuli for a particular setting of the model parameters, where \( s_2 = 1 \) and averaged over \( 10^5 \) trials. Convergence is slower for incompatible stimuli, since the compatibility prior takes time to update from its initial value \( P(M) = 0.9 \).

![Figure 4.2: Mean values of the denominator \( \langle D_t \rangle \) for compatible (blue solid) and incompatible (red dashed) stimuli, each averaged over \( 10^5 \) trials. In both cases the \( \langle D_t \rangle \) rises monotonically toward its upper bound \( 1 + \frac{3}{\sigma^2} = 1.037 \).](image)

Based on these arguments, and in spite of the fact that \( D_t \) can exhibit large variance on individual trials, assume \( D_t \approx \langle D_t \rangle \approx 1 \), and approximate the dynamics of Eq. (4.14) by
the following linear, decoupled system:

\[
\begin{align*}
z_{t+1,1}^t &= \left(1 + \frac{x_1 + x_2 + x_3}{\sigma^2}\right) z_{t+1,1}^{t-1}, \\
z_{t-1,1}^t &= \left(1 + \frac{x_1 - x_2 + x_3}{\sigma^2}\right) z_{t-1,1}^{t-1}, \\
z_{t+1,2}^t &= \left(1 - \frac{x_1 - x_2 + x_3}{\sigma^2}\right) z_{t+1,2}^{t-1}, \\
z_{t-1,2}^t &= \left(1 - \frac{x_1 + x_2 + x_3}{\sigma^2}\right) z_{t-1,2}^{t-1},
\end{align*}
\]  

(4.27)

with initial conditions

\[
z_{0,1,1} = \frac{1}{2}\beta, \quad z_{0,1,2} = \frac{1}{2}(1 - \beta).
\]  

(4.28)

Similar reasoning can be used to derive a linear, decoupled approximation for Eq. (4.19) for the spatial uncertainty model. The approximate dynamics for both models can be written as an iterated linear mapping in the following form

\[
z_{i,j}^t = \left(a_{i,j} + b_{i,j} \eta(t)\right) z_{i,j}^{t-1}, \quad i = 1, \ldots, 4,
\]  

(4.29)

where the random variables \(\eta(t)\) are drawn from a standard normal distribution, and \(a_{i,j}\) and \(b_{i,j}\) are constant parameters whose values depend on the model, the probability being computed, and the compatibility condition of the given trial.

For the compatibility bias model, from the details presented in §4.3.1 and if the current stimulus array \(s(t)\) is compatible and \(s_2 = 1\) I have

\[
a_{i,j} = \begin{cases} 
1 + \frac{3}{\sigma^2}, & i = +1, j = 1, \\
1 - \frac{3}{\sigma^2}, & i = -1, j = 1, \\
1 - \frac{1}{\sigma^2}, & i = +1, j = 2, \\
1 + \frac{1}{\sigma^2}, & i = -1, j = 2,
\end{cases}
\]

and

\[
b_{i,j} = \frac{\sqrt{3}}{\sigma} \quad \forall i, j ;
\]

(4.30)
and if \( s(t) \) is incompatible and \( s_2 = 1 \) I have

\[
a_{i,j} = \begin{cases} 
1 - \frac{1}{\sigma^2}, & i = +1, j = 1, \\
1 + \frac{1}{\sigma^2}, & i = -1, j = 1, \\
1 + \frac{3}{\sigma^2}, & i = +1, j = 2, \\
1 - \frac{3}{\sigma^2}, & i = -1, j = 2,
\end{cases}
\]

and \( b_{i,j} = \frac{\sqrt{3}}{\sigma} \) \( \forall i, j \) . (4.31)

For \( s_2 = -1 \) all the signs in \( a_{i,j} \) above are reversed.

Figure 4.3: Typical distributions from which the multiplicative factors \( a_{i,j} + b_{i,j} \eta(t) \) in Eqn. (4.29) are drawn on each time step. Parameter values are \( \sigma = 1.8 \) (top) and \( a_1 = 0.7, a_2 = 0.3, \sigma_1 = 1.4, \sigma_2 = 1 \) (bottom). For illustrative purposes, standard deviations \( \sigma, \sigma_1, \sigma_2 \) are 20% of those used in the text to reduce overlap of distributions.

For the spatial uncertainty model with compatible stimulus array and \( s_2 = 1 \), the
calculations of §4.3.2 imply:

\[ a_{i,j} = \begin{cases} 
1 - A_1 + 2(a_1 + a_2)(A_2 + A_3) + (a_1 + 2a_2)(A_4 + A_5), & i = +1, j = 1, \\
1 - A_1 - 2(a_1 + a_2)(A_2 + A_3) - (a_1 + 2a_2)(A_4 + A_5), & i = -1, j = 1, \\
1 + A_1 - 2(a_1 + a_2)(A_2 - A_3) + (a_1 + 2a_2)(A_4 - A_5), & i = +1, j = 2, \\
1 + A_1 + 2(a_1 + a_2)(A_2 - A_3) - (a_1 + 2a_2)(A_4 - A_5), & i = -1, j = 2, 
\end{cases} \tag{4.32} \]

and for an incompatible stimulus array and \( s_2 = 1 \):

\[ a_{i,j} = \begin{cases} 
1 - A_1 - 2(a_1 - a_2)(A_2 + A_3) + (a_1 - 2a_2)(A_4 + A_5), & i = +1, j = 1, \\
1 - A_1 + 2(a_1 - a_2)(A_2 + A_3) - (a_1 - 2a_2)(A_4 + A_5), & i = -1, j = 1, \\
1 + A_1 + 2(a_1 - a_2)(A_2 - A_3) + (a_1 - 2a_2)(A_4 - A_5), & i = +1, j = 2, \\
1 + A_1 - 2(a_1 - a_2)(A_2 - A_3) - (a_1 - 2a_2)(A_4 - A_5), & i = -1, j = 2. 
\end{cases} \tag{4.33} \]

In both cases the standard deviation of the noise is given by

\[ b_{i,j} = \begin{cases} 
\sqrt{2(\sigma_1^2 + \sigma_2^2)(A_2 + A_3)^2 + (\sigma_4^2 + 2\sigma_2^2)(A_4 + A_5)^2}, & i = \pm1, j = 1, \\
\sqrt{2(\sigma_1^2 + \sigma_2^2)(A_2 - A_3)^2 + (\sigma_4^2 + 2\sigma_2^2)(A_4 - A_5)^2}, & i = \pm1, j = 2. 
\end{cases} \tag{4.34} \]

Figure 4.3 illustrates normal distributions from which these multiplicative terms in (4.29) are drawn.

### 4.4.2 Taking the continuum limit

I now take continuum limits of the discrete dynamical systems derived above that will allow us to compute properties of interest analytically. First I consider the following finite-difference limit of the iterated mapping (4.29):

\[
\frac{d(z_{i,j}^t)}{dt} = \lim_{\delta t \to 0} \frac{z_{i,j}^{t} - z_{i,j}^{t-\delta t}}{\delta t} = \lim_{\delta t \to 0} \left[ (a_{i,j} - 1) + b_{i,j} \eta(t) \right] z_{i,j}^{t-\delta t}, \tag{4.35} \]

where the \( z_{i,j}^t \) represent the four posteriors \( P(s_2, M|X_t) \). For finite but small \( \delta t = 1/k \), this represents a finer-grained discretization in which \( k \) steps are taken for every one step of (4.29), the deterministic increments being of order \( \delta t \) and the random ones of order
\[ \sqrt{\delta t} \] Taking the limit \( \delta t \to 0 \) in Eq. (4.35), letting \( y_{i,j} = \log(z_{i,j}) \), and appealing to the Ito formula [79, Section 4.1], I obtain independent, uncoupled SDEs for the logarithmic probabilities \( y_{i,j}(t) \):

\[
d y_{i,j} = \left[ (a_{i,j} - 1) - \frac{b_{i,j}^2}{2} \right] dt + b_{i,j} dW \overset{\text{def}}{=} A_{i,j} dt + B_{i,j} dW, \tag{4.36}
\]

with constant coefficients \( A_{i,j} = (a_{i,j} - 1) - \frac{b_{i,j}^2}{2} \) and \( B_{i,j} = b_{i,j} \), whose values are specified in §4.4.1. Since each \( z_{i,j}(t) \) represents a posterior probability, it should take values in the interval \([0, 1]\), so I shall be interested in sample paths \( y_{i,j}(t) \) that start at \( y_{i,j}(0) < 0 \) and satisfy \(-\infty < y_{i,j}(t) \leq 0\).

### 4.4.3 Analytical approximations for the mean posteriors

The SDE (4.36) describes a drift-diffusion process with constant signal and noise level, which has been extensively studied (e.g. [34, 79]). In particular, as noted in §3.1.1, for solutions (sample paths) started at \( y(0) = \mu_0 \) and \( t = 0 \) the probability density function of \( y \) at time \( t \) is the following Gaussian distribution:

\[
p(y,t) = \frac{1}{\sqrt{2\pi\sigma(t)^2}} \exp \left[ -\frac{(y - \mu(t))^2}{2\sigma(t)^2} \right], \tag{4.37}
\]

where

\[
\mu(t) = At + \mu_0 \quad \text{and} \quad \sigma(t) = B\sqrt{t}. \tag{4.38}
\]

(Here and below I drop the subscripts \{i, j\} in \( y \) and \( z \) in the understanding that the appropriate coefficients will be used in the final formulae.) Transforming back into \( z \)-space, using \( y = \log(z) \) and \( dy = \frac{dz}{z} \) yields the density:

\[
p(z,t) = \frac{1}{z\sqrt{2\pi\sigma(t)^2}} \exp \left[ -\frac{(\log(z) - \mu(t))^2}{2\sigma(t)^2} \right]. \tag{4.39}
\]

The inverse transformation \( z = \exp(y) \) takes the Gaussian distribution over \( y \) into a function skewed toward \( z = 1 \), as illustrated in Figure 4.4.

The Gaussian distribution over \( y \) takes positive values on \( y > 0 \) for all \( t > 0 \). This presents a problem, since \( z = \exp(y) > 1 \) for \( y > 0 \), contrary to \( z \)'s designation as a
Figure 4.4: Probability density functions in logarithmic $y$-space and the original $z$-space. Note that the probability distribution in $y$-space is a Gaussian distribution, while transferred into $z$-space, the distribution is skewed into one with squeezed head and extended tail.

...probability measure. Therefore, when computing expected values of $P(s_2, M|X_t)$, which requires integration of the quantity $zp(z, t)$, all values of $z > 1$ will be replaced by $z = 1$ (or values of $y > 0$ by $y = 0$ in the equivalent integral over $y$). However, to retain analytical tractability, I continue to assume a Gaussian distribution over $y$ at time $t$ when generating the distribution at time $t+1$ – that is, I only replace the inappropriate values of $y$ (or $x$) in the integral, not in the underlying drift-diffusion process. The expected (mean) value of $z$ is therefore approximated as

$$
\langle z(t) \rangle \approx \int_0^1 z p(z, t) \, dz = \int_0^1 \frac{1}{\sqrt{2\pi\sigma(t)^2}} \exp\left[ -\frac{(\log(z) - \mu(t))^2}{2\sigma(t)^2} \right] z \, dz \\
+ \int_1^\infty \frac{1}{\sqrt{2\pi\sigma(t)^2}} \exp\left[ -\frac{(\log(z) - \mu(t))^2}{2\sigma(t)^2} \right] dz,
$$

(4.40)

which may be evaluated as explained in Appendix B to yield

$$
\frac{\exp\left[\mu(t) + \frac{\sigma(t)^2}{2}\right]}{2} \left[ 1 - \text{erf}\left(\frac{\mu(t) + \sigma(t)^2}{\sqrt{2\sigma(t)^2}}\right) \right] + \frac{1}{2} \left[ 1 + \text{erf}\left(\frac{\mu(t)}{\sqrt{2\sigma(t)^2}}\right) \right].
$$

(4.41)

Substituting values appropriate for the compatibility bias model from Eqs. (4.30-4.31) for the parameters $a_{i,j}$ and $b_{i,j}$, and hence for $A_{i,j}, B_{i,j}$, and via Eqs. (4.38), for $\mu(t)$ and
\[ \sigma(t), \text{ yields estimates for the four mean posterior probabilities at time } t: \]

\[
\langle P(s_2, M|X_t) \rangle \approx \frac{1}{2D(t)} \times \\
\left\{ \exp \left[ \mu(t) + \frac{(\sigma(t))^2}{2} \right] \left[ 1 - \text{erf} \left( \frac{\mu(t) + (\sigma(t))^2}{\sqrt{2\sigma(t)^2}} \right) \right] + \left[ 1 + \text{erf} \left( \frac{\mu(t)}{\sqrt{2\sigma(t)^2}} \right) \right] \right\}. \tag{4.42}
\]

where \( D(t) \) is the sum of all four probabilities that normalizes the expressions, and for compatible stimuli the functions \( \mu(t) \) and \( \sigma(t) \) are:

\[
\mu(t) = \begin{cases} 
+ \frac{3t}{2\sigma^2} + \log \left( \frac{\beta}{\sqrt{2}} \right), & s_2 = +1, M = 1, \\
- \frac{9t}{2\sigma^2} + \log \left( \frac{\beta}{2} \right), & s_2 = -1, M = 1, \\
- \frac{5t}{2\sigma^2} + \log \left( \frac{1-\beta}{2} \right), & s_2 = +1, M = 2, \\
- \frac{t}{2\sigma^2} + \log \left( \frac{1-\beta}{2} \right), & s_2 = -1, M = 2,
\end{cases} \tag{4.43}
\]

and for incompatible stimuli:

\[
\mu(t) = \begin{cases} 
- \frac{5t}{2\sigma^2} + \log \left( \frac{\beta}{\sqrt{2}} \right), & s_2 = +1, M = 1, \\
- \frac{t}{2\sigma^2} + \log \left( \frac{\beta}{\sqrt{2}} \right), & s_2 = -1, M = 1, \\
+ \frac{3t}{2\sigma^2} + \log \left( \frac{1-\beta}{2} \right), & s_2 = +1, M = 2, \\
- \frac{9t}{2\sigma^2} + \log \left( \frac{1-\beta}{2} \right), & s_2 = -1, M = 2,
\end{cases} \tag{4.44}
\]

Here, the facts that all sample paths start with the initial conditions specified in Eq. (4.10) and that \( \mu_0 = \mu(0) = \log(z(0)) \) are also used.

As noted at the beginning of this section, normalization and averaging do not commute.

This may be understood in terms of the distributions of Figure 4.4 as follows. While each sample path can be computed for the uncoupled processes and normalized at time \( t \) to yield the same result as a sample path of the coupled system (cf. Figure 4.8), different normalization factors must typically be applied to the values of different paths \( z_{i,j}(t) \) at each time \( t \). This would distort the distributions \( p(z_{i,j}, t) \), thereby changing their means. However, I may appeal to the observation that the expected value of the denominator remains close to 1 (cf. Figure 4.2) to conclude that this distortion is likely to be small, and
proceed by dividing by the sums of the four mean probability trajectory values at time $t$ to normalize the resulting expressions.

Typical results for mean posterior probabilities are shown in Figure 4.5. The approximate predictions developed above are shown as dashed curves and the results of averaging over 5000 simulated trials of the full inference model (4.9) are shown solid; compatible and incompatible trials are shown in red and blue respectively. As above, I compute 200 steps for the discrete iteration of the full system, and I evaluate the corresponding quantities for $t \in [0, 200]$ time units from the formula above. For $P(M) = 0.5$ (not shown), joint posteriors for correct responses increase similarly for both compatible and incompatible cases, but $P(M) = 0.9$ elicits markedly different behaviors (top left). The compatibility posteriors $P(M = 1|X_t)$ show a general rise for compatible stimuli and a monotonic fall for incompatible stimuli, but the posterior probability $P(s_2 = 1|X_t)$ shows a significant dip below 0.5 at early times for incompatible stimuli, while it rises monotonically for compatible stimuli. As discussed below and in Chapter 6, the resulting accuracies exhibit similar patterns to the experimental data, with the incompatible case showing a dip in accuracy for early responses. Evolutions of the four individual posterior probabilities are shown in the lower panels of Figure 4.5.

Figure 4.5 illustrates that, while the approximations developed here do not capture all the detailed behavior of the full model, they do provide reasonably good approximations to the average evolutions of the posteriors over the course of a trial. Time scales are slightly misestimated and the compatibility posterior $P(M = 1|X_t)$ (top right) fails to reproduce the slight dip below 0.9 that occurs for compatible trials at early times, but the relative orderings of all the posteriors are correctly predicted. Overall, absolute errors in mean posteriors, computed as described at the end of this section, lie between 0.002 and 0.05, the largest being for $P(M = 1|X_t)$ in the case of incompatible stimuli (top right, lower curves).

Predictions for the spatial uncertainty model follow from the formula (4.42) in a similar manner, upon the substitution of values for $a$ and $b$ from Eqs. (4.32-4.34), and using the initial conditions $\mu_0 = \log(1/4)$ for all four posteriors (Eq. (4.21)). For compatible stimuli,
the function $\mu(t)$ is

$$
\mu(t) = \begin{cases} 
\frac{a^2 + a^2}{\sigma_1 + \sigma_2} - \frac{2a_1 a_2}{\sigma_1 + 2\sigma_2} t + \log\left(\frac{1}{4}\right), & s_2 = +1, M = 1, \\
\frac{3a_1^2 + 8a_1 a_2 + 3a_2^2}{\sigma_1 + \sigma_2^2} - \frac{2(a_1 + a_2)(a_1 + 4a_2)}{\sigma_1 + 2\sigma_2} t + \log\left(\frac{1}{4}\right), & s_2 = -1, M = 1, \\
\frac{-3a_1^2 + 4a_1 a_2 + a_2^2}{\sigma_1 + \sigma_2} + \frac{2(3a_1 - 4a_2)a_2}{\sigma_1 + 2\sigma_2} t + \log\left(\frac{1}{4}\right), & s_2 = +1, M = 2, \\
\frac{a_1^2 + a_2^2}{\sigma_1 + \sigma_2} - \frac{2a_1 (a_1 - 3a_2)}{\sigma_1 + 2\sigma_2} t + \log\left(\frac{1}{4}\right), & s_2 = -1, M = 2;
\end{cases}
$$

(4.45)

for incompatible stimuli

$$
\mu(t) = \begin{cases} 
\frac{-3a_1^2 - 4a_1 a_2 + a_2^2}{\sigma_1 + \sigma_2} - \frac{2(3a_1 + 4a_2)a_2}{\sigma_1 + 2\sigma_2} t + \log\left(\frac{1}{4}\right), & s_2 = +1, M = 1, \\
\frac{a_1^2 - 4a_1 a_2 - 3a_2^2}{\sigma_1 + \sigma_2} - \frac{2a_1 (a_1 + 3a_2)}{\sigma_1 + 2\sigma_2} t + \log\left(\frac{1}{4}\right), & s_2 = -1, M = 1, \\
\frac{a_1^2 + a_2^2}{\sigma_1 + \sigma_2} + \frac{2a_1 (a_2 - 3a_2)}{\sigma_1 + 2\sigma_2} t + \log\left(\frac{1}{4}\right), & s_2 = +1, M = 2, \\
\frac{3a_1^2 - 8a_1 a_2 + 3a_2^2}{\sigma_1 + \sigma_2} - \frac{2(a_1 - a_2)(a_1 - 4a_2)}{\sigma_1 + 2\sigma_2} t + \log\left(\frac{1}{4}\right), & s_2 = -1, M = 2;
\end{cases}
$$

(4.46)

and in both cases

$$
\sigma(t) = \begin{cases} 
\sqrt{\left(\frac{(a_1 + a_2)^2}{\sigma_1^2 + \sigma_2^2} + \frac{(a_1 + 2a_2)^2}{\sigma_1^2 + 2\sigma_2^2}\right) t}, & s_2 = \pm 1, M = 1, \\
\sqrt{\left(\frac{(a_1 - a_2)^2}{\sigma_1^2 + \sigma_2^2} + \frac{(a_1 - 2a_2)^2}{\sigma_1^2 + 2\sigma_2^2}\right) t}, & s_2 = \pm 1, M = 2.
\end{cases}
$$

(4.47)

The above results, presented in Figure 4.6, are not as good as those for the compatibility bias model. Nonetheless, the approximate model captures the key features of the evolving posteriors in the full model rather well, predicting the relative ordering of the posteriors appropriately in all cases except the incorrect choices $P(<<>)$ and $P(>>>)$ for incompatible stimuli; in that case the approximation for $P(>>>)$ diverges from the correct function, increasing rather than decreasing as $t$ increases (lower right panel), for an absolute error of 0.12. Apart from this case, however, errors lie between 0.015 and 0.08.

The errors for both models were computed for each mean posterior using the $L^1$ norm as follows:

$$
\text{Error} = \sum_{t=0}^{T} |p_t - \tilde{p}_t|,
$$

(4.48)

where $p_t$ and $\tilde{p}_t$ denote the posteriors predicted by the full and simplified models respectively.
4.4.4 Making use of explicit mean posteriors

In addition to providing explicit expressions for posterior probabilities, the continuum limit also yields approximations for accuracy and reaction time distributions. To estimate accuracy as a function of response time under the free response protocol assumed by [117], I compute the fraction of mass of the evolving probability density \( p(z_{t,1}, z_{t,2}) \) that exceeds a given threshold \( z_{t,1} + z_{t,2} = q \) at each time \( t \) (recall Eq. (4.11)). This is essentially the same as the interrogation protocol calculated in §3.1.1 above. This procedure overestimates first passage times, since some of the sample paths that lie beyond the threshold \( q \) at time \( t \) may have crossed at earlier times, but it permits some analytical simplification. Without loss of generality, I shall assume that \( s_2 = 1 \).

The integral to be evaluated is

\[
P(s_2=1|X_t)_{\text{est}} = \int_0^q \int_{q-z_2}^q p(z_{t,1}, z_{t,2}) \, dz_{t,1} \, dz_{t,2} \approx \int_0^q \int_{q-z_2}^q p(z_{t,1}) p(z_{t,2}) \, dz_{t,1} \, dz_{t,2},
\]

(4.49)

where the shorthand notation \( p(z_j, t) = p(z_{t,j}) \) is used, and the approximation comes from assuming \( p(z_{t,1}, z_{t,2}) \approx p(z_{t,1}) p(z_{t,2}) \) for the uncoupled and linearized approximate dynamical system – this assumption greatly simplifies the computations, although the uncoupled processes are not entirely independent since they are activated by common inputs \((x_1, x_2, x_3)\), albeit in different linear combinations. Also note that the variables \( z_j \) should be non-negative (cf. Figure 4.4). The domain of integration is pictured in Figure B.1. The \( p(z_j, t)'s \) take the forms derived in §4.4.3 above and since each is a normalized Gaussian in the logarithmic \( y \) variables, the integral of their product over the entire positive quadrant is 1. Hence I have

\[
P(s_2=1|X_t)_{\text{est}} = 1 - \int_0^q \int_0^{q-z_2} p(z_{t,1}) p(z_{t,2}, t) \, dz_{t,1} \, dz_{t,2},
\]

(4.50)

which is evaluated in the Appendix B to yield:

\[
P(s_2=1|X_t)_{\text{est}} = \frac{3}{4} - \frac{1}{4} \text{erf} \left( \frac{\log(q) - \mu_2(t)}{\sqrt{2} \sigma_2(t)} \right) - \frac{1}{2} \int_0^q p(z_{t,2}, t) \text{erf} \left( \frac{\log(q - z_2) - \mu_1(t)}{\sqrt{2} \sigma_1(t)} \right) \, dz_2,
\]

(4.51)
where
\[ p(z_2, t) = \frac{1}{z_2 \sqrt{2\pi \sigma_2(t)^2}} \exp \left[ -\frac{(\log(z_2) - \mu_2(t))^2}{2\sigma_2(t)^2} \right]. \] (4.52)

Unfortunately, the final integral in Eq. (4.51) cannot be computed analytically, but it can be evaluated accurately and rapidly by numerical methods.

Response accuracy is approximated by the fraction of correct responses that exceed threshold:
\[ \frac{P(s_2 = 1|X_t)_{\text{est}}}{P(s_2 = 1|X_t)_{\text{est}} + P(s_2 = 2|X_t)_{\text{est}}}, \] (4.53)
where the denominator approximates the sum of all four probabilities \( z_{1,1}^T + z_{1,2}^T + z_{2,1}^T + z_{2,2}^T \). (The term \( P(s_2 = 2|X_t)_{\text{est}} \) is computed in a similar manner to Eq. (4.51), with the appropriate expressions for \( \mu(t), \sigma(t) \) from §4.4.3.) The denominator is the cumulative reaction time and so its derivative with respect to \( t \) provides the reaction time distribution. Hence, both accuracy and reaction time distributions can be approximated semi-analytically. Figure 4.7 shows the resulting approximations to the mean posteriors for the compatibility bias model, for a particular setting of model parameters. The dip in accuracy for incompatible trials is reproduced, and after an initial rise in accuracy for compatible trials, accuracy slowly declines. Also, reaction time distributions are shifted rightward for incompatible trials compared to compatible trials, as in the data of Figure 4.7.

As noted above, sample paths of the SDE (4.36) may pass across \( q \) and back, possibly repeatedly, in the interval \((0, t)\), so these results do not directly correspond to the first-passage decision policy of the Bayesian models in [117]. This accounts for differences between the accuracy curves and reaction time distributions of Figure 4.7 and the free response results of §4.5. For example, the compatibility bias free response data of Figure 4.9 do not show the mild decline in accuracy for later compatible trials of Figure 4.7, although the spatial uncertainty simulations of Figure 4.10 do show such a decline. Nonetheless, the qualitative agreement between Figures 4.7 and 4.9 is quite good, and since the semi-explicit expression Eqs. (4.51-4.52) replaces lengthy Monte-Carlo simulations of §4.5, it may be helpful in guiding parameter fits to data.

The posterior probability expressions can also be used to constrain parameter choices,
by requiring the derivative of $P(s_2 = 1|X_t) = z_{1,1}^t + z_{1,2}^t$ at time $t = 0$ to be negative and finding corresponding conditions on the parameters. The results of this computation (details not shown) agree closely with those in Section 4.3.

### 4.5 Direct simulations and comparisons

Direct simulations of the linear approximation can be compared with those of the original inference model. Figure 4.8 shows the results for the compatibility bias model for a particular setting of parameters ($\sigma = 9$), comparing the full inference model with the simplified iteration of (4.16). The same sequence of noisy observations $x_i(t)$ was used for both processes and in computing the value of $P(s_2 = 1|X_t)$ for the latter at each timestep $t$, normalization was applied only at that step. The agreement is remarkably good, validating my linear approximations to the products of probabilities (4.5-4.4) developed in Section 4.3. The quality of the linear approximation for the spatial uncertainty model is similarly good (details not shown).

I can also simulate perceptual discrimination based on the linearized evidence accumulation process, using the first passage criterion for threshold crossing appropriate for free response conditions. As in [117], I adopt the decision threshold $q = 0.9$ for both the compatibility bias and the spatial uncertainty model. The time span, taken here as 200 steps, is divided into ten bins and sample paths for the full model (4.9) and the approximate decoupled system (4.16) and its analogue for spatial uncertainty are computed. The decoupled results are then normalized by dividing by the sum $\sum_{i,j} v_{i,j}^t$ at each $t$ in the current bin (normalization is not applied for steps 1 through $t-1$). The same (unit) step size is used in all cases. Responses are logged when the first of the probabilities $P(s_2 = +1|X_t) = P(s_2 = +1, M = 1|X_t) + P(s_2 = +1, M = 2|X_t)$ or $P(s_2 = -1|X_t) = P(s_2 = -1, M = 1|X_t) + P(s_2 = -1, M = 2|X_t)$ crosses $q$. After collecting sufficiently many paths (2000 in this case), response time histograms are formed and the fraction of correct responses in each bin summed to yield accuracy vs. time curves.

Figures 4.9 and 4.10 illustrate the results of such simulations for the compatibility and
spatial uncertainty models respectively. Accuracy vs. reaction time, and empirical distributions of reaction time are shown for both the full and approximate models. The approximate systems reproduce the characteristic dip in accuracy for fast incompatible trials for both models, and the accuracy curves and reaction time distributions predicted by the approximate theory agree well with those of the full inference models. Note that the use of the first passage criterion for response produces reaction time distributions that agree with the exact model in details of their shapes: a rise at short reactions times to a peak, followed by a long tail. The distributions for incompatible trials are also flatter and shifted rightward compared to those for compatible trials, as in the data of Figure 4.7.

4.5.1 Fitting the models to data

I now briefly describe the results of fitting the full models of Section 4.2 and the reduced DD processes of Sections 4.4.2-4.4.4 to the data both of [43] and [92], reproduced in Fig. 4.7. For the compatibility bias model the parameters fitted are the noise level $\sigma$, prior $\beta$, threshold $q$ and step durations $\delta t$ (for DDM) and $\Delta T$ (for the full model), which determine the overall timescale. For spatial uncertainty, they are $\sigma_1$, $\sigma_2$, $a_1$, $q$ and $\delta t$, $\Delta T$ (as in §4.3.2, set $a_2 = 1 - a_1$). As in §3.2, the Matlab routine fmincon() was used with the same weighted Euclidean error norm as in Eqn (3.22). (see §3.2 for details). The parameter values obtained to fit the data of [43]: $\sigma = 6.5$, $\beta = 0.90$, $q = 0.9$, $\delta t = 0.43$ ms , $\Delta T = 2.3$ ms for the compatibility bias model and $\sigma_1 = 6.4$, $\sigma_2 = 5.5$, $a_1 = 0.72$, $q = 0.83$, $\delta t = 0.9$ ms and $\Delta T = 1.1$ ms for the spatial uncertainty model. The parameter values obtained to fit the data [92] are: $\sigma = 6.7$, $\beta = 0.86$, $q = 0.95$, $\delta t = 0.58$ ms, $\Delta T = 1.6$ ms for the compatibility bias model, and $\sigma_1 = 6.9$, $\sigma_2 = 5.2$, $a_1 = 0.71$, $q = 0.92$, $\delta t = 2$ ms and $\Delta T = 0.4$ ms for the spatial uncertainty model. The parameters are consistent with the assumptions of Sections 4.3 and 4.4.1-4.4.2: e.g., $1/\sigma^4 \ll 1/\sigma^2$ (cf. equation (4.13)). Fitting errors are as follows: fitting to [43] data: compatibility bias: DDM 4.6; full model 5.4; spatial uncertainty: DDM 4.1; full model 4.0. Fitting to [92] data: compatibility bias: DDM 4.2; full model 4.4; spatial uncertainty: DDM 3.8; full model 4.5.
The results are shown in Figure 4.11 and Figure 4.12. In Figure 4.12, fit qualities are similar for both full models (they differ by less than 2.3%), but for the DDM reductions the spatial uncertainty model wins by 10%. In both cases, surprisingly, the approximate DDM fits somewhat better than the full Bayesian procedure. Both models underestimate RTs and fail to capture the drop in accuracy at short RTs on compatible trials (left panels). They do reproduce this drop on incompatible trials, although the full compatibility bias model does not exhibit the dip below 50%, and both it and its DDM reduction underestimate RTs. The spatial uncertainty model is substantially better in this regard (lower right panel). In Figure 4.11, I tried a modified norm that preferentially weights low RT data: this slightly improved fits of RT distributions, but did not affect compatible accuracy fits. I also fitted the full and DD models to the data of [43] (Fig. 1.1(A)), obtaining similar fit qualities, although the failure to capture the steady rise from 50% accuracy at low RTs for compatible trials is more striking in that case.

The fits are not as good as those obtained in Chapter 3 for the data of [43] using the DDM with variable drift rates. However, there the error of 2.4 is obtained with 8 free parameters, compared with 4 and 5 respectively in the present cases. Indeed, in §3.1-3.2, 6 parameters are used to describe the drift rates in the compatible and incompatible cases, modeling progressive increase in attention to the central stimulus, and these cases are fitted separately. In the present study compatible and incompatible trials are fit simultaneously.

I also remark that there is substantial variability in Eriksen data, perhaps due to differing deadlining protocols. (Deadlines are necessary to produce enough short reaction times and hence obtain a significant dip in accuracy on incompatible trials.) The resulting variability in motor preparation times can affect reaction times, and no allowance for this is made in the inference models per se, which describe only cognitive processing.

### 4.6 Discussion

In Chapters 2-3, a neural network model of the Eriksen task [23, 92] was linearized and reduced to a DDM with time-varying drift, allowing relatively complete analysis that reveals
how parameters influence accuracy curves such as those of Figure 4.7. However, this network model involves somewhat arbitrary assumptions on architecture and parameters, and it is not clear how the DDM reduction of Chapter 3, with its variable drift rate, relates to the optimal decision theory for the constant drift case [11]. The present Chapter addresses this issue by offering analytically tractable approximations to the optimal Bayesian inference model proposed in [117].

Specifically, the joint signal probability distribution of Eqn. 4.4 is approximated as a linear sum, and then, by assuming that the sum of the non-normalized posteriors remains close to one and taking a continuum limit, analytical expressions are obtained for the mean posterior probabilities. Employing a further approximation in which the net probabilities of having answered correctly or incorrectly at time $t$ are computed, semi-analytical approximations for accuracy and reaction time distributions are then derived. While the latter correspond more closely to an “interrogation protocol” [11, 66] in which subjects are cued to respond at specific times, and so differ quantitatively from those computed numerically for free responses (compare Figure 4.7 with Figure 4.9), the overall accuracy curves and individual posteriors derived from the continuum model reproduce those of the Bayesian model quite well (see Figures 4.5-4.6). These analytical approximations are therefore expected to be useful in guiding parameter selection when fitting models to experimental data.

In Section 4.3, provides an example of this by deriving simple parametric constraints that must hold to obtain the dip below 50% in the posterior probability for early responses.

Moreover, although the coefficients differ, the linearized update rules of both Eqns. (4.14) and (4.19) demonstrate that the flanker inputs $x_1$ and $x_3$ work with the target input $x_2$ for the compatible hypotheses, and against it for the incompatible hypotheses. This underlying computational architecture gives rise to the same basic ability of both the compatibility bias and spatial uncertainty models to account for the dynamics of flanker interference in behavioral data.

This analysis also reveals that a particularly simple stochastic differential equation, the constant-drift diffusion (DD) process of Eqn. (4.36), approximately describes the evolution
of Bayesian posteriors in log probability space. As described in [11], this is a continuum limit of the sequential probability ratio test [108], which is known to be optimal for identifying noisy signals in two-alternative choice tasks [109]. Moreover, as shown in Chapters 2-3 and in [11, 66], DD and related Ornstein-Uhlenbeck processes emerge naturally in linearized reductions of competing leaky accumulator models [106] for 2AFC. In these neural networks the difference between activities in a pair of units at the output decision or response stage behaves like the accumulating variable $y(t)$ in Eqn. (4.36) (In N-alternative choice models, linear combinations of variables approximate $(N-1)$-dimensional DD processes [106, 72].) [39]. DD models can also capture bottom-up (stimulus-driven) and top-down influences such as attention and expectation of rewards via variable drift rates [66, 27].

Since accumulator models may be derived from biophysical models of spiking neurons [110, 115], in which their activities represent short-term averages of collective firing rates, this suggests a mechanism by which neural substrates may be able to perform Bayesian computations. Specifically, in reducing the coupled Bayesian inference model (Eqn. (4.9)) to a DD process I see how prior information maps into initial conditions, and evolving posteriors in log probability space are represented by spike rates of groups of neurons. In connection with the latter, I note that there is [12] present computational and experimental evidence that Bayesian computations involving exponentiation and taking logarithms, as in Section 4.4, can be approximated by neurons in the basal ganglia [118, 12].
Figure 4.5: Predictions of the full and simplified compatibility bias models in the case that the central symbol is S ($s_2 = 1$) and with prior compatibility bias $P(M) = 0.9$. Top left: marginal mean posterior probabilities $P(s_2 = 1|M)$ (correct response) for compatible and incompatible conditions. Top right: marginal mean posterior $P(M = 1)$ for compatibility. Bottom row: individual mean posteriors for compatible (left) and incompatible (right) trials. Results from full inference model, averaged as in Figure 4.5, shown solid and predictions of the continuum approximation (4.42-4.44) shown dashed. Keys identify individual curves.
Figure 4.6: Predictions of the full and simplified spatial uncertainty models. Top left: marginal mean posterior probabilities $P(s_2 = 1|M)$ (correct response) for compatible and incompatible conditions. Top right: marginal mean posterior $P(M = 1)$ for compatibility. Bottom row: individual mean posteriors for compatible (left) and incompatible (right) trials. Results from full inference model, averaged as in Figure 4.5, shown solid and predictions of the continuum approximation (4.42) and (4.45-4.47) shown dashed. Keys identify individual curves.
Figure 4.7: Predictions of accuracy (left) and reaction time histograms (right) computed under the approximation of Section 4.4.4. Solid curve and dark boxes indicate compatible trials; dashed curve and light boxes indicate incompatible trials.

Figure 4.8: Posterior probability $P(s_2 = 1|X_t)$ for one sample path of the approximate compatibility bias model (Eq. (4.16), dashed), compared with a sample path from the original inference model (Eq. (4.9), solid). The same sequence $x(t)$ of observations was used in both cases.
Figure 4.9: Accuracy and reaction time distributions for the compatibility bias model for compatible stimuli (left) and incompatible stimuli (right). Solid and right hand (blue) bar of each RT bin pair from full inference model of [117]; dashed and left hand (yellow) bars from approximate (linearized likelihood) model. Results obtained by averaging over 2,000 simulated trials.

Figure 4.10: Accuracy and reaction time distributions for the spatial uncertainty model for compatible stimuli (left) and incompatible stimuli (right). Format as in Figure 4.9. Results obtained by averaging over 2,000 simulated trials.
Figure 4.11: Accuracy (upper curves) and reaction time distributions (lower curves) from the full (squares) and reduced DD (triangles) models for compatible (left) and incompatible (right) trials. Upper panels show compatibility bias and lower panels spatial uncertainty model results respectively. Parameters were fitted to the data of [43] (dashed curves).
Figure 4.12: Accuracy (upper curves) and reaction time distributions (lower curves) from the full (squares) and reduced DD (triangles) models for compatible (left) and incompatible (right) trials. Upper panels show compatibility bias and lower panels spatial uncertainty model results respectively. Parameters were fitted to the data of [92] (dashed curves).
Chapter 5

A Model for Sequential Effects in the Eriksen task

5.1 Experimental data of sequential effects in Eriksen task

Sequential effects in the Eriksen task have been reported in [44]. Figure 5.1 shows average reaction times and error rates in human subjects under different experimental conditions. These data indicate that responses to compatible trials are faster if the previous trial was also compatible (by 10 ms) and that responses to incompatible trials are faster (although by only 3 ms) and more accurate (error rate was 0.075 lower) if the previous trial was also incompatible. In other words, faster and more accurate responses are given on trials in which the “level of distraction” is repeated (at least for incompatible trials).

For convenience, I use notions which consist of a consequence of symbol “C”s and “I”s to indicate the compatibility of trials, and the order of the symbols indicates the order of trials. For example, “CI” trial represents that the previous trial is compatible and the current trial is incompatible, and so on. Note there is a controversy in the experimental history of the Eriksen task. One study shows that “CI” trials were slower than “II” trials for eight of the eleven subjects [14], and some other studies show that “CI” trials were slower than “II” trials only for response repetition trials [77].
Figure 5.1: Average reaction times (left panel) and error rates (right panel) for each of four experimental conditions. Responses to compatible trials were faster if the previous trial was also compatible (by 10 ms) and responses to incompatible trials were faster (although by only 3 ms) and more accurate (error rate was 0.075 lower) if the previous trial was also incompatible. Left: mean reaction times over all trials. Right: error rates. From [44] Figure 1.

Conditional accuracy functions for compatible and incompatible trials, sorted on the basis of the compatibility level on the previous trial, are shown in Figure 5.2. These data reveal that the sequential effect is larger after a compatible trial than after an incompatible trial. There are significant primary effects of reaction time bin (greater accuracy for slow responses than for fast responses), and of distraction (greater accuracy for compatible trials than for incompatible trials), as well as a Reaction Time Bin × distraction interaction (the difference in accuracy between compatible and incompatible trials was more evident for fast responses than for slow responses). More interestingly, there is also a significant two-way interaction between the level of distraction on the previous and current trials: The noise effect was larger after a compatible trial than after an incompatible trial [44]. Another
interesting aspect of the data shown in Figure 5.2 is that accuracy for incompatible trials is below chance for responses given between 200 and 250 ms after stimulus onset, which is consistent with the data shown in Figure 1.1.

Figure 5.2: Conditional accuracy functions for the four experimental conditions. (Response latency is shown in milliseconds.) Experiment reveals that the sequential effect was larger after a compatible trial than after an incompatible trial. From [44] Figure 2.

Several studies have been done to model and simulate sequential effects in the Eriksen task. For example, in [13], a conflict monitoring unit was introduced into the neural network, proposed in [23] to simulate the Eriksen task. It implements a feedback loop connecting conflict monitoring to cognitive control. The model accounts for the data well, and the theory that conflict monitoring is reflected in the activity is consistent with a large body of existing data. Moreover, an fMRI study subsequently provided direct evidence, showing that the anterior cingulate cortex (ACC), on the medial surface of the frontal lobes of the brain, serves to detect and signal the occurrence of conflicts in information processing [14].

However, several problems remain. For example, in [14], it is claimed that psychophysi-
ological evidence indicates that conflict monitoring may act as a source of feedback to mechanisms involved in recruiting attention, serving to indicate the need for increased top-down control on information processing. But the original model neither evaluates this potential relationship between conflict monitoring and cognitive control, nor investigates the relation between the ACC and other brain areas implicated in executive function. Moreover, the conflict monitoring mechanism is unable to predict the long term sequential effect, i.e. whether the neural system has adapted to the task in long term. Finally, the conflict monitoring model of [14] did not explore the sub-optimality of human subject behaviors in the Eriksen task. In this chapter, I address these questions and propose possible answers.

5.2 A model for sequential effects of Eriksen task

I base my model on the Bayesian inference compatibility bias model of §4.2. This model involves encoding the prior knowledge that spatially proximate stimuli usually have similar sensory properties. As such, flanker stimuli information may be over-processed at the start of a trial, but later, as evidence accumulates, the salient central stimulus may eventually overwhelm the prior expectations. There is a built-in mechanism which is asymmetrical on being compatible and incompatible. Since the behavioral evidence of sequential effects described above derives from a series of stimuli with different compatibilities, the compatibility bias model appears more suitable than the spatial uncertainty model.

Recall that the basic assumption for the compatibility bias model is that the probability of being compatible is biased, e.g. the prior \( P(M = 1) = 0.9 \). As noted above, in the natural environment, nearby stimuli tend to have similar properties, and there is electrophysiological and psychophysical evidence (e.g. Gestalt laws) showing that the visual system generally encodes this property. It is therefore likely that, without extensive training on the Eriksen task, subjects may have a bias greater than 0.5 in assuming a particular trial to be compatible. However, when actually engaged in the Eriksen task, the compatibility of the previous trial could temporarily weigh more than all previous experiences. The prior may therefore change on a trial-to-trial basis according to the compatibility of previous trials.
In my model, I increase the compatibility bias prior if the previous trial is compatible, and decrease it if the previous trial is incompatible, adapting its value according to the following rules:

If the previous trial is compatible, the prior for the current trial is increased by

\[ \beta_{n+1} = \beta_n + \delta (1 - \beta_n); \]  

(5.1)

If the previous trial is incompatible, the prior for the current trial is decreased by

\[ \beta_{n+1} = \beta_n - \delta \beta_n. \]  

(5.2)

Here \( n \) is the trial number, and \( \delta \in (0, 1) \) is a small increment, set to 0.05 in the simulations below. Note that equations (5.1) - (5.2) maintain the constraint that the prior bias for each trial \( \beta_n \) is restricted to the interval \([0, 1]\).

Figure 5.3: Given equal probability of being compatible and incompatible, the prior gradually decays to 0.5. The simulation is averaged over 1000 trials, with initial biased prior \( \beta = 0.9 \), and \( \delta = 0.05 \).

Note that if compatible trials and incompatible trials are equally probable, the compatibility prior slowly decays to 0.5 (Figure 5.3), i.e. compatibility bias disappears after
a sufficiently long period of time. The system gradually changes from asymmetrical to symmetrical. Since experimental protocol requires equal probability of being compatible and incompatible, I shall observe dropping priors, but with a short term memory effect. In section 5.3.4 I discuss this issue, and show supporting evidence from experimental data analysis.

5.3  Comparison with empirical data

In this section, I perform simulations of the full Bayesian inference model of Chapter 4 argumented by the simple adaptive prior introduced in §5.2. At first, only the influence from the previous trial is considered. Then a simulation of different response latency bins is given. Subsequently, the influence from longer history is included. Finally, a comparison of prior updating with empirical data is shown.

5.3.1  Sequential effects of immediately previous trial

The data used here were originally collected by N. Yeung, and analyzed in [116]. The experimental method is detailed in Appendix C. All behavioral data are averaged over 12 subjects, 816 trials each. All 816 trials are divided into 12 blocks, with 68 trials in each block. The inter-trial interval is about 1000 – 1100ms. There were long time intervals between blocks, during which subjects could rest and review their performance.

In Figure 5.4, the top panels show the experimental data, which contain averaged reaction times for correct trials (left) and error rates (right) for each of the experimental conditions. The data presented in Figure 5.4 indicate that responses to compatible trials are faster if the previous trial was also compatible (by 8 ms) and that responses to incompatible trials are faster (by only 4 ms) and more accurate (the error rate was 0.075% lower) if the previous trial was also incompatible. In other words, faster and more accurate responses were given on trials in which the compatibility was repeated (at least for incompatible trials). These findings are consistent with the previous study [44], see Figure 5.1.
In the simulation, I assume that adaptation of the prior occurs only when the inter-trial interval is short. Once one block is finished and subjects have a long period of rest, I assume that the prior resets to its original value. The resulting simulations describe the experimental data very well. “CC” trials are faster than “CI” trials by about 4 time steps and “II” trials were faster than “CI” trial by about 2 time steps and also more accurate (the error rate was 0.06 lower). The simulation parameters for the Bayesian inference model are: $\sigma = 6.3$, $\delta = 0.05$, $q = 0.92$, with initial prior bias $\beta_1 = 0.9$, which are the same as Chapter 4.
This numerical experiment supports the assumptions that the compatibility prior is biased, and can be changed by immediate compatibility history. The compatibility of the previous trial can help subjects make decisions faster and more accurately, no matter whether the actual central stimulus sequence be a repetition or an alternation. Note that the units of time axes on RT plots are the same for experimental data and simulation results. The simulation is transformed using the formula \( RT_{\text{experiment}} = T_0 + c \times RT_{\text{simulation}} \), where \( T_0 \) represents the time required for initial visual processing and motor response, and \( c \) has unit ms/time step, showing how much time one simulation step represents. In this simulation, \( T_0 = 227 \) ms, \( c = 3.8 \) ms/time step.

5.3.2 Sequential effects in different response latency bins

In this section, conditional accuracy functions for each of the experimental sequential conditions are considered. Figure 5.5 is simulated under the same experimental conditions as the data of 5.2 reproduced in Figure 5.2. In [44], it is shown that the dip is more evident after a compatible trial. In fact, although the accuracy for responses to incompatible trials with latencies between 200 and 249 ms was significantly below chance after a compatible trial, it was not significantly below chance after an incompatible trial. This finding indicates that, after a compatible trial, greater weight is given to a process that is sensitive to distraction.

From the simulation, I plot the conditional accuracy functions for each of the experimental sequence conditions as a function of response time. The resulting curves match the experimental findings qualitatively in that a compatible trial makes the dip deeper. The simulation parameters are: \( \sigma = 8.3, \delta = 0.05, q = 0.9 \), and initial prior bias \( \beta_1 = 0.9 \).

5.3.3 Sequential effects of longer response histories

In this section, I plot reaction time and accuracy from an Eriksen task experiment as a function of preceding stimulus compatibility sequences. The data used here were also collected by N. Yeung [116]. The experimental setup is the same as in Section 5.3.1, and details of the procedure can be found in Appendix C. Reaction times plotted below are
Figure 5.5: Simulation of conditional accuracy functions for each of the experimental conditions. (response latency is shown in 20 step bins. Compare with Figure 5.2).

Figure 5.6 shows both simulated (bottom panels) and empirical data (top panels). Initial inspection of the behavioral data reveals that there is no significant difference for accuracy if the current trial is compatible, no matter what the sequence history is. This is also true for mean reaction times of correct trials when the current trial is incompatible. The simulations show similar results for both situations.

Figure 5.6 also demonstrates that if the current trial is incompatible, more incompatible trials in the recent history imply more accurate responses on the current trial. It is interesting that the mean reaction time remains similar no matter what compatibility combinations occurred in the response history. Similar effects appear if the current trial is incompatible. In contrast, the current trial is compatible, more compatible trials in recent history imply faster responses, but no significant difference in accuracy. The simulation reproduces these results. The simulation parameters are the same as in §5.3.1: \( \sigma = 6.3, \delta = 0.05, q = 0.92, \) with initial prior bias \( \beta_1 = 0.9. \) In this and next section, I use \( T_0 = 237 \text{ ms}, c = 3.6 \text{ ms/time} \)
Figure 5.6: Reaction time and accuracy as a function of preceding stimulus compatibility sequences. Top panels are experimental data, while bottom ones are simulation results. “C-I” sequences read downward, so that the current trial appears at the bottom in all cases.

5.3.4 Experimental evidence for prior updating

As I discussed above, the prior adaptation rule implies that the prior value drops from its original value gradually if compatible and incompatible stimuli have equal probability of being presented. Yeung’s data [116] show that this happens in his experiment as well.

To do this, I separate data from each block into two parts. Since the inter-trial-interval
within a block is fairly short ≈ 1000ms, I assume the prior continually adapts according to Eqns. (5.1) and (5.2). Between blocks, the prior is reset to its original value. Therefore, if I separate each block into a first half and a second half, the mean priors should be different, and consequently, the mean RT and accuracy should also differ between the first and second halves. The experimental data is shown below in Figure 5.7 (top panels). The blue curves are averaged values over trials in first half, and red curves are for the second half. Red curves in RT plots are slightly higher than blue curves in call cases. In the accuracy plot, there is no significant difference between the curves if the current trial is compatible, but the accuracy is higher for second half blocks if the current trial is incompatible.

The simulation (bottom panels) reproduces most of the features apparent in the experimental data. Here the simulation parameters are: \( \sigma = 6.3, \delta = 0.05, q = 0.92 \). Starting prior values for the first half (blue) and second half (red) are \( \beta_1 = 0.9 \) and \( \beta_2 = 0.85 \) respectively.

Another possible mechanism that might shift the curves is to change thresholds. Previous studies [99] and analyzing the present data show that subjects can adapt their threshold slowly to a lower value during a block of trial. The direct effect of lowering the threshold is to lower the RT curves, but not raise them. However, in the data analysis of two halves data in one block, I found the curves were shifted up. Therefore, it is impossible to achieve the results by only lowering the threshold, and prior updating is necessary to derive the effects in Figure 5.7.

## 5.4 Discussion

In this chapter, I propose a simple extension to the Bayesian inference model of Chapter 4 to simulate sequential effects in the Eriksen task. In this model, the prior compatibility bias adapts according to the compatibility/incompatibility of the previous trial. Model simulations agree with the experimental findings, and the experimental data also provide evidence for prior adaptation in continuous trials.

In Chapter 4, I did dynamical system analysis on a compatibility bias model and a
Figure 5.7: Reaction time and accuracy as a function of preceding stimulus compatibility sequences. Blue curves: averaged value over trials in first half of all blocks. Red curves: averaged value over trials in second half of all blocks. Top panels are experimental data, while bottom ones are simulation results.

Spatial uncertainty model, both proposed to model the Eriksen task. Each model has its own generative features and fundamental assumptions. The compatibility bias model assumes that the brain may be wired, through evolution and/or development, to encode the prior knowledge that spatially proximate stimuli have similar sensory properties. The hypothesis made in the spatial uncertainty model is that receptive fields of visual cortical neurons are partially overlapping, resulting in flanker stimuli interference at the beginning of trials. The compatibility bias and spatial uncertainty models therefore make different
predictions in certain respects, including sequential effects. Repeated training under the equal prior condition should gradually eliminate the “CI” dip, evident in Figures 5.2 and 5.5, and push performance toward the pattern observed from the equal prior model. Previously, this has been observed briefly in monkeys trained repeated under the Eriksen paradigm (Gary Aston-Jones). The present work on human behavioral data confirms this prediction, and further proposes a model of sequential effects which fits data well.

Traditional explanations of sequential effects have been divided broadly into two types: facilitation and expectancy [68]. Facilitation is thought to involve a low-level mechanism [60], while expectancy effects are attributed to higher-level strategic processes [107], which mainly influence task performance at longer RSIs. Previous studies of sequential effects (e.g. [20]) generally use connectionist models, which are biologically plausible and extensively studied. They tease apart different cognitive mechanisms via separation of timescales and different RSIs, as well as shifting thresholds and starting points [20, 33]. In Chapter 4, in which I reduce the full Bayesian inference model to a continuous drift diffusion process (Eqn. 4.36), the starting points of the DDM can be written as the logarithm of the prior probability ratios. Therefore, updating priors is equivalent to updating starting points of the connectionist models. This work not only further confirms basic assumptions of the Bayesian inference model, but also connects it with traditional views of sequential effects.

Sequential effects in two-alternative forced choice tasks have been studied for a long time. Previous work has explored various combinations of repetition and alternation detection schemes in an attempt to account for empirical results [20]. In this chapter, combinations of compatibility have been extensively considered instead of repetition an alternation due to the special experimental scheme of Eriksen Task. A recent experimental exploration of sequential effects of combinations of repetition and alternation in the Eriksen task [77] finds that mean RTs and error rates for each combination of trial type and previous trial type show a clear reduction of the compatible effect following incompatible trials. The authors claim that this pattern of results provides strong evidence that the conflict-adaptation effect reflects associative stimulus-response priming instead of conflict-driven adaptation in
cognitive control. In my model, the prior probabilities of repetition and alternation are implicitly assumed equal, and not updated during trials. In future work, a mechanism for updating these prior probabilities could be included in the model and tested.
Chapter 6

Conclusions

Here I summarize the work described above and briefly note open problems and proposals for future directions.

6.1 Summary of the dissertation

In Chapters 2 and 3 of this dissertation I analyze a linearized version of the connectionist model for the Eriksen two-alternative forced-choice flanker task presented in [23] and [92]. I show that, analytical solutions of a decoupled, linearized model can provide reasonable estimates of critical times and hence reproduce the characteristic dip in accuracy plot. I also show that the dynamics of the two-unit decision layer can be decoupled and reduced to a drift-diffusion model (DDM) with time-varying drift rate. I consider both the interrogation protocol and free response protocol, and show that the DDM produces acceptable approximations to accuracy and reaction time distributions. It also provides a slightly better fit to empirical data than does the full model, while using fewer parameters.

In Chapter 4, I review the Bayesian inference models proposed to simulate and analyze the Eriksen task in [117], and derived simplified models: uncoupled, linear discrete dynamical systems. From these I derive the continuum limit, a stochastic ordinary differential equation (SDE), which, in logarithmic probability space, is also a DDM. As described in [11], this is also the continuum limit of the SPRT, which is known to be optimal for...
identifying noisy signals in TAFC tasks. From this, I compute analytical predictions for mean trajectories of the posterior probabilities. I also compute accuracy vs. time curves and reaction time distributions and fit the empirical data.

The work summarized above exemplifies how both connectionist models and normative Bayesian inference models of cognitive tasks can be related to the DDM, which then provides an analytic framework for interpreting empirical observations. On the one hand, I have shown that a DDM reduction of a multi-layer model of the Eriksen task suggests that processing involves a progressive change in the drift rate over the course of a trial, reflecting the influence of top-down attentional mechanisms. This finding can be related, in turn, to an optimality analysis that generates new hypotheses about the factors governing attentional control (e.g., prior expectations of stimulus compatibility). On the other hand, in many neural networks both bottom-up (stimulus-driven) and top-down influences are modeled, and evidence accumulation occurs. Since accumulator models may in turn be derived from biophysical models of spiking neurons [110, 115], this suggests a mechanism through which neural substrates may be able to approximate Bayesian computations, which are also optimal under certain conditions. I built connections between these two difference mechanisms for optimality analysis, and for the first time, showed that evolving posteriors are represented by spike rates of groups of neurons by reducing the coupled Bayesian inference model to a DDM.

The results of Chapter 4 are extended applied in Chapter 5. Based on the Bayesian inference model described and studied in Chapter 4, I proposed a simple mechanism for trial by trial updating of priors that can account for sequential effects in the Eriksen task. Simulations of the model agree with experimental findings, and experimental data also provide evidence for prior updating from trial to trial. In Chapter 4, I did dynamical system analysis on a compatibility bias model and a spatial uncertainty model, both proposed to model the Eriksen task. Each model has its own generative features and fundamental assumptions although they produced similar results in RT and accuracy plots. The compatibility bias model assumes that the brain may be wired, through evolution and/or development, to
encode the prior knowledge that spatially proximate stimuli have similar sensory properties. The work sequential effects described in Chapter 5 supports this assumption, and the resulting model for sequential effects fits data well.

6.2 Open problems and future directions

Neural networks and normative Bayesian inference models exemplify are two distinct approaches in the study of computational neuroscience. They both have advantages as well as shortcomings. Neural network models are biologically based, and some of their mechanisms have supporting experimental evidence, such as the Hebbian learning rule. However, one can create a “successful” neural net without understanding how it worked, due to non-linearity and high dimensionality. On the other hand, the Bayesian inference model is built on statistical inference in which evidence or observations are used to update or to newly infer the probability that a hypothesis may be true. Although its simplicity and optimality is alluring, biologically realistic implementations remain elusive. In Chapter 4 of this dissertation, a possible connection between neural network models and Bayesian inference models was demonstrated under appropriate conditions. This work suggests that the two approaches are not mutually exclusive, but rather may work together in the Eriksen task. One of the possible future directions is to continue this study in a more general way.

As noted in Chapter 4-5, when I reduce full Bayesian inference model to a continuous drift diffusion process (Eqn. (4.36)), the starting points of the DDM can be written as the logarithm of the prior probabilities. Therefore, updating priors is equivalent to updating starting points of the connectionist models. Further analysis could be done to connect sequential effect in Bayesian inference model and DDM, for example in the sequential effects model of Chapter 5.

In Chapter 5, I mentioned that a conflict monitoring unit was introduced to study sequential effects in the Eriksen task [23]. In previous studies conflict monitoring by the anterior cingulate cortex (ACC) was posited to signal a need for greater cognitive control, producing neural and behavioral adjustments. In a recent fMRI study, it is shown that ACC
conflict-related activity predicts both greater prefrontal cortex activity and adjustments in behavior, supporting a role for ACC conflict monitoring in the engagement of cognitive control [59]. But none of the study has produced insights into the specific mechanisms by which detection of conflict in ACC engages the recruitment of control. Noted in Chapter 4-5 and [14], both prior probabilities and ACC activities were updated from trial to trial during Eriksen task, and prior probabilities were also related to neural activities from top-down feedbacks. The above problem could possibly solved by implementing ACC conflict-related activity with Bayesian probabilities.

This dissertation focuses on a specific 2AFC paradigm: the Eriksen task. However, many of the ideas and models presented here could be generalized to other cognitive tasks. For example, similar in some ways to the Eriksen task, the Stroop task also produces conflict between responses [105]. In this task names of colors are presented to subjects as words displayed in different colored inks. Participants must name the ink color of a color word while ignoring the meaning of the word. To perform the task subjects need to select the relevant information (the ink color) and suppress irrelevant information (the meaning of the word). The Stroop task provides evidence for the automaticity of reading. Participants are usually slower and less accurate in identifying color words printed in an incongruent compared to a congruent ink color. Theoretical work has been done to study the Stroop task, and a computational model was proposed in [21], which is similar to the neural network model for Eriksen task. An extension of the current work could include the dynamical system analysis, normative modeling, and sequential effect studies of the Stroop task, and indeed, of other attention-related decision tasks.
Appendix A

Solution of a Six-dimensional Differential Equation

I observe that an $n \times n$ matrix $A$ with diagonal entries $-k$ and off-diagonal entries $gw$ has two eigenvalues

$$
\lambda_1 = -(k + (n-1)gw), \quad \lambda_2 = gw - k, \quad (A.1)
$$

with multiplicities 1 and $n-1$ respectively, and since $A$ is symmetric, $n-1$ mutually orthogonal eigenvectors belonging to $\lambda_2$ can be found. Along with the eigenvector $(1, 1, \ldots, 1)$ of $\lambda_1$, these yield an orthonormal transformation $T$ with $T^{-1} = T^T$ with the latter given explicitly by:

$$
y_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i; \quad y_j = \sqrt{\frac{j-1}{j}} \left[ \frac{1}{j-1} \sum_{i=1}^{j-1} x_i - x_j \right], \quad 2 \leq j \leq n. \quad (A.2)
$$

For $n = 6$ the orthonormal eigenvector matrix is:

$$
T = \begin{bmatrix}
1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{12} & 1/\sqrt{20} & 1/\sqrt{30} \\
1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{12} & 1/\sqrt{20} & 1/\sqrt{30} \\
1/\sqrt{6} & 0 & -2/\sqrt{6} & 1/\sqrt{12} & 1/\sqrt{20} & 1/\sqrt{30} \\
1/\sqrt{6} & 0 & 0 & -3/\sqrt{12} & 1/\sqrt{20} & 1/\sqrt{30} \\
1/\sqrt{6} & 0 & 0 & 0 & -4/\sqrt{20} & 1/\sqrt{30} \\
1/\sqrt{6} & 0 & 0 & 0 & 0 & -5/\sqrt{30}
\end{bmatrix}, \quad (A.3)
$$

87
and (2.8) transforms to the uncoupled system $\dot{y} = T^T A T y + T^T I$:

\begin{align*}
\dot{y}_1 &= \lambda_1 y_1 + \frac{(a+2b)\sqrt{6}}{6}, \\
\dot{y}_2 &= \lambda_2 y_2 + \frac{b}{\sqrt{2}}/ - \frac{b}{\sqrt{2}}, \\
\dot{y}_3 &= \lambda_2 y_3 + \frac{(b-2a)\sqrt{6}}{6}, \\
\dot{y}_4 &= \lambda_2 y_4 + \frac{(a+b)\sqrt{12}}{12}, \\
\dot{y}_5 &= \lambda_2 y_5 + \frac{a-3b\sqrt{20}}{\sqrt{20}}/ + \frac{a+b\sqrt{20}}{\sqrt{20}}, \\
\dot{y}_6 &= \lambda_2 y_6 + \frac{(a+2b)\sqrt{30}}{30} + \frac{(a-4b)\sqrt{30}}{30},
\end{align*}

(A.4)

where the first of each alternative additive term corresponds to compatible stimuli and the second to incompatible stimuli, and the single terms in components 1, 3 and 4 apply to both.

The initial value problem

\[ \dot{y} = \lambda y + f(t), \ y(0) = y_0 \]

(A.5)

has the solution:

\[ y(t) = y_0 e^{\lambda t} + \int_0^t e^{\lambda(t-s)} f(s) \, ds, \]

(A.6)

and in case that $y_0 = 0$ and $f(t) = At + B$, I have

\[ y(t) = \frac{1}{\lambda^2} \left[ A(e^{\lambda t} - 1 - \lambda t) + \lambda B(e^{\lambda t} - 1) \right]. \]

(A.7)

In the case of balanced parameters ($k = w \Rightarrow \lambda_2 = 0$), (A.7) becomes:

\[ y(t) = \frac{A t^2}{2} + Bt. \]

(A.8)

Equipped with these solutions of (A.5), I compute $p = T y(t)$ and sum the appropriate components to obtain Equations (2.11-2.17) of §2.3. In doing so I also appeal to fact that the 4-dimensional subspace $p_1 = p_5, p_2 = p_6$ is invariant, implying, via $y = T^T y$, that solutions started at $y_j(0) = 0$ satisfy

\[ \frac{5 y_5}{\sqrt{20}} = - \left( \frac{y_2}{\sqrt{2}} + \frac{y_3}{\sqrt{6}} + \frac{y_4}{\sqrt{12}} \right), \quad \frac{6 y_6}{\sqrt{30}} = \frac{1}{5} \left( \frac{6 y_2}{\sqrt{2}} - \frac{4 y_3}{\sqrt{6}} - \frac{4 y_4}{\sqrt{12}} \right). \]

(A.9)

88
Appendix B

Evaluation of an Integral

To evaluate the integrals of Eq. (4.40) I employ the change of variables

\[ x = \frac{(\log(z) - \mu)}{\sqrt{2\sigma^2}}, \quad z = \exp\left(\mu + \sqrt{2\sigma^2}x\right), \quad (B.1) \]

so that \( dx = \frac{dz}{z\sqrt{2\sigma^2}} \) and the integrals become

\[ \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-x^2 + \mu + \sqrt{2\sigma^2}x} \, dx \quad \text{and} \quad \frac{1}{\sqrt{\pi}} \int_{\infty}^{-\infty} e^{-x^2} \, dx. \quad (B.2) \]

The second expression is a standard error function integral, and the first may be put into the same form by completing the square in the argument of the exponent:

\[ x^2 - \mu - \sqrt{2\sigma^2}x = \left(x - \sqrt{\frac{\sigma^2}{2}}\right)^2 - \left(\mu + \frac{\sigma^2}{2}\right), \quad (B.3) \]

followed by the further change of variables

\[ u = \left(x - \sqrt{\frac{\sigma^2}{2}}\right). \quad (B.4) \]

This process results in the expressions of Eq. (4.41).

To evaluate the integral of Eq. (4.50) I proceed as follows, dropping the explicit reference to time dependence, which enters the expressions through the mean and standard deviations
Figure B.1: The integral of the joint posterior probability distribution is taken over the positive \((z_1, z_2)\)-quadrant less the shaded triangular region.

\(\mu(t), \sigma(t)\). Figure B.1 indicates the domain of integration.

\[
P(s_2 = i|\mathbf{X}_t)_{\text{est}} = 1 - \int_0^q \int_0^{q-z_2} p(z_1) p(z_2) \, dz_1 \, dz_2 = 1 - \int_0^q p(z_2, t) \int_0^{q-z_2} p(z_1) \, dz_1 \, dz_2
\]

\[
= 1 - \int_0^q p(z_2, t) \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{\log(q - z_2) - \mu_1(t)}{\sqrt{2}\sigma_1(t)^2} \right) \right] \, dz_2
\]

\[
= 1 - \frac{1}{2} \int_0^q p(z_2) \, dz_2 - \frac{1}{2} \int_0^q p(z_2) \, \text{erf} \left( \frac{\log(q - z_2) - \mu_1(t)}{\sqrt{2}\sigma_1(t)^2} \right) \, dz_2
\]

\[
= 1 - \frac{1}{4} \left[ 1 + \text{erf} \left( \frac{\log(q) - \mu_2(t)}{\sqrt{2}\sigma_2(t)^2} \right) \right] - \frac{1}{2} \int_0^q p(z_2) \, \text{erf} \left( \frac{\log(q - z_2) - \mu_1(t)}{\sqrt{2}\sigma_1(t)^2} \right) \, dz_2. \quad \text{(B.5)}
\]

Here I have added subscripts to the time-varying means and standard deviations \(\mu_j(t), \sigma_j(t)\), using the same shorthand \(z_j = z_{1,j}\) as in \(\S 4.4.4\) to indicate which of the four cases \(s_2 = \pm 1; M = 1, 2\) enumerated in \(\S 4.4.3\) is intended.
Appendix C

Experimental Method

Experimental Method ([116])

Participants. Nine female and seven male undergraduate students from Princeton University (Princeton, NJ) participated in the Eriksen task experiment. It is a single 2-hr session, which can be used for a psychology course credit. All students were right-handed and between 18 and 23 years old, and all had normal or corrected-to-normal vision. Informed consent was obtained from each participant at the start of the session.

Procedure. The participants were seated in front of a screen in a dimly lit room. They performed a version of the flanker task in which they responded by key-press to indicate the direction of a central arrow that was surrounded by flanker arrows. There were four stimuli, the congruent stimuli <<<<< and >>>>> and the incongruent stimuli <<<<< and >>>>>. The four stimuli were presented in pseudo-random order with the constraint that each stimulus appeared equally often in each block. On each trial, the participant was first presented with a fixation cross in the center of the screen. The cross was replaced after 500 ms with an imperative stimulus. The stimulus was presented for 100 ms and then the screen was cleared, remaining clear until 500 ms after the participants response. At this time the string ”–” appeared to mark the intertrial interval, the duration of which varied randomly from 1,000 to 1,100 ms. All stimuli were presented in white on a black background. At a viewing distance of roughly 110 cm, the arrow stimuli each subtended
0.4° of visual angle vertically and 0.6° horizontally, and they were spaced 0.3° apart.

Participants first performed 2 or 3 practice blocks of 36 trials each. They then performed 12 blocks of 68 trials each during which behavioral and ERP data were collected. The participants were allowed to rest between blocks, at which time they were given feedback showing their mean correct RT and error rate in the previous block and for the whole session. If their error rate fell below 8%, they were encouraged to respond more quickly. If they made more than 16% errors, they were told to respond more carefully. Participants were also encouraged through verbal instruction to sit in a relaxed position, to minimize eye movement, and to blink as seldom as possible while they performed the task.
References


